

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 1, 2003, 22:34:30 ; Search time 5590 Seconds

(without alignments)
4386.849 Million cell updates/sec

Title: US-10-030-330-1

Sequence: 1 MKKSFLLAVMFGIAMQGH.....PNGAYILKVDGTTKINIVH 843

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

ched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_plus.model -DEV=xlh
-Q=gsn2.1/USPTO_spool/US10030330/runat_23052003_181644_6751/app_query.fasta.1.1031
-DB=GenEmbl -QMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cd1 -LIST=45
-DOCAI=200 -THR_SCORE=0.0 -THR_MAX=100 -THR_MIN=0 -ALIGN=20 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030330.ecgn.1.1.3750 -runat_23052003_181644_6751 -NCTP=6 -ITP=3
-NO_XIPXY -NO_MAP -LARGEOUT=NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb_da: *
2: gb_hg: *
3: gb_in: *
4: gb_ov: *
5: gb_ov: *
6: gb_ov: *
7: gb_ov: *
8: gb_ov: *
9: gb_ov: *
10: gb_ov: *
11: gb_ov: *
12: gb_ov: *
13: gb_ov: *
14: gb_ov: *
15: gb_ov: *
16: gb_ov: *
17: gb_ov: *
18: gb_ov: *
19: gb_ov: *
20: gb_ov: *
21: gb_ov: *
22: gb_ov: *
23: gb_ov: *
24: gb_ov: *
25: gb_ov: *
26: gb_ov: *
27: gb_ov: *
28: gb_ov: *

29: em_v1: *
30: em_hg_hum: *
31: em_hg_inv: *
32: em_hg_other: *
33: em_hg_mus: *
34: em_hg_pln: *
35: em_hg_rod: *
36: em_hg_mam: *
37: em_hg_vrt: *
38: em_sy: *
39: em_hgo_hum: *
40: em_hgo_mus: *
41: em_hgo_other: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4438	100.0	2532	6	AX048260	AX048260 Sequence
2	1240	27.9	2661	1	POYAAA	M83096 Porphyromon
3	1238	27.9	4056	1	S75942	S75942 prt-99 kda
4	439	11.2	1357	1	STRSPBJ	L26134 Streptococc
5	439	11.2	1357	1	STRSPBJ	L26134 Streptococc
6	439	11.2	1357	1	STRSPBJ	L26134 Streptococc
7	439	11.2	1357	1	STRSPBJ	L26134 Streptococc
8	438	11.2	1479	1	AB051298	L26157 Streptococc
9	436	11.2	1357	1	STRSPBJ	L26154 Streptococc
10	436	11.2	1357	1	STRSPBJ	L26154 Streptococc
11	436	11.2	1357	1	STRSPBJ	L26154 Streptococc
12	436	11.2	1357	1	STRSPBJ	L26154 Streptococc
13	436	11.2	1357	1	STRSPBJ	L26154 Streptococc
14	436	11.2	1357	1	STRSPBJ	L26154 Streptococc
15	436	11.2	1357	1	STRSPBJ	L26154 Streptococc
16	436	11.2	1357	1	STRSPBJ	L26154 Streptococc
17	436	11.2	1357	1	STRSPBJ	L26154 Streptococc
18	435	11.2	1357	1	STRSPBJ	L26154 Streptococc
19	435	11.2	1357	1	STRSPBJ	L26154 Streptococc
20	435	11.2	1357	1	STRSPBJ	L26154 Streptococc
21	435	11.2	1357	1	STRSPBJ	L26154 Streptococc
22	435	11.2	1357	1	STRSPBJ	L26154 Streptococc
23	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
24	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
25	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
26	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
27	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
28	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
29	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
30	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
31	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
32	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
33	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
34	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
35	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
36	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
37	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
38	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
39	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
40	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
41	434	11.1	1357	1	STRSPBJ	L26154 Streptococc
42	433	11.1	1357	1	STRSPBJ	L26154 Streptococc
43	433	11.1	1357	1	STRSPBJ	L26154 Streptococc
44	433	11.1	1357	1	STRSPBJ	L26154 Streptococc
45	433	11.1	1357	1	STRSPBJ	L26154 Streptococc

RESULT 1

ALIGNMENTS

AX048260
 LOCUS AX048260 2532 bp DNA linear PAT 15-DEC-2000
 DEFINITION Sequence 2 from Patent WO0063394.
 ACCESSION AX048260
 VERSION AX048260.1 GI:11877029

KEYWORDS
 SOURCE
 ORGANISM

Porphyromonas gingivalis.
 Porphyromonas gingivalis
 Bacteria; Bacteroidetes; Bacteroides; Bacteroidales;
 Porphyromonadaceae; Porphyromonas.
 1 (bases 1 to 2532)
 Travis, J., Potempa, J., and Nelson, D.
 A polypeptide having amidolytic activity for a serpin
 Patent: WO 0063394-A 2 26-OCT-2000;
 UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. (US) ; Travis,
 James (US) ; Potempa, Jan (US) ; Nelson, Daniel (US)

FEATURES
 Location/Qualifiers

1..2532
 /organism="Porphyromonas gingivalis"
 /db_xref="taxon:837"

Alignment Scores:

Pred. No.: 5 25e-283 Length: 2532
 Score: 4438.00 Matches: 843
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

US-10-030-330-1 (1-843) x AX048260 (1-2532)

QY 1 MetLysSerPheLeuLeuAlaIleValMetLeuPheGlyIleAlaMetGlnGlyHis 20
 DB 1 ATGAAAAAAGTTTCTTTTACCATAGTAATGCTTTGGCATTCGCCATGACGACAT 60
 QY 21 SerAlaProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnVal 40
 DB 61 TCTGCTCCGGTTACGAAAGAGCGACTTTGAGTCTGGCTCGGCTGGCTTGGACAGGTA 120
 QY 41 SerLeuArgMetGlyGlnThrAlaValSerAspLysIleSerIleAspTyrValIleArg 60
 DB 121 TCCCTTGGCAATGGCAACACAGCATCTGCACAAAGATTTCATCGATTACGTTATTCGG 180
 QY 61 GlnGlyAspAlaGlyArgGlyIleThrSerGlnGlnGlySerProAlaTyrPheTyr 80
 DB 181 CACAGAGATGCTGACAGAGGGGTTCACATCACAAGGAAGGCTCTCCGCAATTTTAT 240
 QY 81 ValAlaAsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIlePro 100
 DB 241 GTAGCTAATCGTGAATATGAGGGCTATGCTTGTAGCAGCAGATGACAGAAATACCG 300
 QY 101 ThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAspAsnLeu 120
 DB 301 ACAATTTTAGCCTATTTCACCCATTTGGCTTGCACATGACAGATATGCCGGAACAATCTT 360
 QY 121 ArgMetTyrLeuGlnIleTyrAspGlnGlnIleGlyLeuIleLeuSerGlyLysAlaGln 140
 DB 361 CGCATGTGGCTCAAAATTTACGATCAGAAATAGGCTCATCTTCCGGAAGACTCG 420
 QY 141 LeuAsnGlnGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMetAsp 160
 DB 421 CTCATATGAAGATATTACGTACGACGAGGGCGTACCGGCTGAAGTACATCTCTGATGAT 480
 QY 161 AsnGlyHisPheAlaAsnAspPrometArgTyrPheGlnGlnGlyTyrProTyrAsnLys 180
 DB 481 AACGGTCAATTTTGGCAACGATCCCATGGCATGAAATCAAGGTTACCCCAAGCAATATAG 540
 QY 181 GluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAlaThrAlaAla 200
 DB 541 GAACCACTGCTTCTTAATGCAATCATGCCCTATACCGGCTGTGTGCTACTCTGACGCA 600

QY 201 GlnIleMetArgTyrHisSerTyrProLeuGlnGlnGlySerPheAspTyrHisAla 220
 DB 601 CAATATCATGCGCTACCATAGCTGGCGCTCAAGGTGAAGGCTCTTGCATATATCATCA 660
 QY 221 GlySerLeuValGlyAsnTyrPheGlyGlnMetTyrAspTyrIleAsnMet 240
 DB 661 GGTTCATATAGTTGGCACTGCTCCGCAATTTGGTGAATGTACACTGTGATCAATATG 720
 QY 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAlaPheTyrAlaThrLeu 260
 DB 721 CCGGAAATCCCGACCTTATATCTGATATCTGATCAATGAGGATGAGCTTACGCCACATG 780
 QY 261 MetArgAspValSerIleSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
 DB 781 ATGCGTATGATGAGCATCTGTTTCCATGATGATTTTATGAAAAAGAGGTGTCTGAC 840
 QY 281 SerValTyrValAlaGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeu 300
 DB 841 AGCGTTATATGATGAGAGCTTGGCAACAACTTCCCTACACCGCTTCACTGACGCTA 900
 QY 301 HisValArgAlaLeuTyrThrSerGlnGlnTyrPheAspMetIleArgGlyLeuAla 320
 DB 901 CATGACGCGCTTATATACCTACACAGAGGTGCACAGATATATGTCGGGGGAATCTGCC 960
 QY 321 SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValLys 340
 DB 961 TCCGGAAGCCGCTATATATGACAGGATATACAGACATAGACATAGCTTTCCTTGGC 1020
 QY 341 AspGlyTyrAlaSerAspGlyThrPheHisPheAsnTyrGlyTyrGlyValSerAsn 360
 DB 1021 GATGGTTATGCTCCGATGATCTTCCATTTCACTGGGGTGGGAGGTCTTCCAAC 1080
 QY 361 GlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGlnGlyLe 380
 DB 1081 GGCCTCTACAACTACACACTCTCTCCGCACTTGTGGGTATGAGAGTACGAGGATA 1140
 QY 1141 GGTATTACCATTTATCAAGATCATCACCGGTATCGAACCGGCTAAACTCCCTGAA 1200
 QY 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluAlaGlyTyrLysSer 420
 DB 1201 GCCGTACAGATGCTTCCGCACTTGGCACTGCAAGAAACATGAAAGCCGAGATATAAGT 1260
 QY 421 GluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlnGlnSerAsnLeu 440
 DB 1261 GATCCGATTTGAAGTGAAGGTATTCATATATATATACAGTGAAGACAAATCAATCTT 1320
 QY 441 AspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluValIleGluValLysThrSerSer 460
 DB 1321 GACCTCGGATACAGATTGACAAAGGCTGACGAGAAAGTCATAGAGGTGAAAACCTTCACT 1380
 QY 461 IleAsnIleSerTyrTyrGlyTyrGlnHisProGlyUserPheSerLeuAlaProAsn 480
 DB 1381 ATCAATATCTCTTGGTATGATGATGAGAGCATCCCGAGAGTGTCTCATTTGGCACTTAT 1440
 QY 481 GlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGlnGln 500
 DB 1441 CAGTTGTCAAGAGATCAACACCATCACTTATTCGTCGACAGGCAACGACAG 1500
 QY 501 TrpGluProValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsnThrThr 520
 DB 1501 TGGAGCCGCTACGCGATGCAAGAGAGATTTGTCATATAGCATTAAGTAAATCGCA 1560
 QY 521 AspProAsnAsnValValAlaThrAlaAspAsnAsnGlnGlyLysLeuSerIleValPro 540
 DB 1561 GACCCGAACATGTGTGATGTCAGGTATATTAACGAGGCACTCAATTCGTCCCC 1620
 QY 541 AsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPheAsn 560
 DB 1621 AACAGCTTGTGCGAGATCTGATTTATGAACTATAGACATATACAGTATCAAT 1680
 QY 561 SerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGlyAlaThr 580


```

Db      1681 AGGACAGCCCTATAGATCCGTACACCCGTTGCTTCTATCTACAGACGCTACT 1740
Qy      581 AAlaspaSpVal1leSerleuglytrpValMetAlaGluValProGlyGlySerSeran 600
Db      1741 GCGGACGATGATATCTTTGGCTGGTAAATGGCTGAAGTTCCGGGGGTACAGACAAAC 1800
Qy      601 TyrProVal1trSerlysaSpVal1leuThrleuSerGluLysAspTyrThrleuTrp 620
Db      1801 TATCCGGTGGTGGTCTTAAAGACGTTCTACCTCTCGGAGACGCACTATACATTTGTGG 1860
Qy      621 TyrArgPheSer1leasnaSnglnLysAspGluTrpLysLysIleGlySerValSerVal 640
Db      1861 TATAGATTTTCCATCAACAACAAAGATGATGAAAAAGTCCGGAACCGTGCAGTA 1920
Qy      641 LysThrProThrGluTyrThrHisProLeuPheGluValGlyHisSnglnThrSerThr 660
Db      1921 AAAACACCAAGAGATACGACCCCTTATTCGAAAGTGGCCATATCAACTTCTAC 1980
Qy      661 TyrThrleuAspMetAlaHisAsnArgVal1leuProAspPheThrleuLysAsnLeuGly 680
Db      1981 TATACCTGGATATGGACACACACAGATATGCCCCGACTTACACTCAAAAATTCGGA 2040
Qy      681 LeuProPheAsnGlyLysLeuValVal1PheArgGlnThrGlnSerSerGlySer 700
Db      2041 TTGCCCTTCAATGGTGAAGTGGTGTGTTTCCGCCAAACACAACTTCATGGGGTCT 2100
Qy      701 LeuThrAlaAlaGlnGluThrValHisIleLysGlnGlyGluThrPheValTyrLysPro 720
Db      2101 TTATGGGCGACCTCAACAAACAGTACATATCAACGAGAGAACTTCGATATTAACCT 2160
Qy      721 Val1Val1GluGlyProIleProAspGlySerTyrArgAlaThrleuHisAlaPheValAsn 740
Db      2161 GTTGTCGAAGGCCCTTACCTGATGATGATCTTATCGGACCCCTCATGATTCGTAAC 2220
Qy      741 GlyGlnGlnGlnLeuTyrLeuLysGlyLysArgAsnTyrThrValLysIleValAsnGly 760
Db      2221 GGACACACACAGTGTGACTCAAGGGGAAAGAAAGAACTACACGGTGAAGATCGTCAATGG 2280
Qy      761 ThrAlaVal1GluAlaIleLeuSerSerGluGluIleArgVal1PheProAsnProAlaArg 780
Db      2281 ACAGCGGTGAAGCAATAGATGATCGTGAAGATGATATTCCTTAATCCGGACCG 2340
Qy      781 AspTyrVal1GluIleSerAlaProCysIleProGlnGluThrSerIleIleuPheAsp 800
Db      2341 GATATGTGGAATATCGGACCTTGCATTCCTCCCAAGAAACATCTATCTTTGAT 2400
Qy      801 LeuSerGlyLysIleValMetLysAsnSerleuSerAlaGlyHisGlyArgMetAspVal 820
Db      2401 CTGTGCGCAAGATGTCTAGAAATAGTTATATACGGGGGCGATGCGAATGATGTC 2460
Qy      821 SerArgLeuProAsnGlyLysAlaTyrIleLeuLysValAspGlyTyrThrThrLysIleAsn 840
Db      2461 ACCCGACTTCTTAATGGGCGCTACATCTTAAGGTGATGATATCGACGAAATAAAT 2520
Qy      841 IleValHis 843
Db      2521 ATAGTCAC 2529

```

```

RESULT 2
LOCUS      POYAAAA                      2661 bp    DNA          linear    BCT 22-MAR-1994
DEFINITION Porphyromonas gingivalis gene sequence.
ACCESSION  M83096
VERSION     M83096.1 GI:463172
KEYWORDS
SOURCE
ORGANISM   Porphyromonas gingivalis DNA.
            Porphyromonas gingivalis
            Bacteria; Bacteroidetes; Bacteroides; Bacteroidales;
            Porphyromonadaceae; Porphyromonas.
REFERENCE  1 (bases 1 to 2661)
            Ootogo,J. and Kuramitsu,H.K.
            Isolation and characterization of the Porphyromonas gingivalis prtI
            TITLE

```

```

gene, coding for protease activity
JOURNAL    Infect. Immun. 61 (1), 117-123 (1993)
MEDLINE    9314862
PUBMED     8093357
COMMENT     On Mar 24, 1994 this sequence version replaced gi:150840.
FEATURES
    source   1..2661
             /organism="Porphyromonas gingivalis"
             /db_xref="taxon:837"
BASE COUNT  697 a 662 c 646 g 655 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1.3e-72      Length: 2661
Score:      1240.00     Matches: 299
Percent Simlarity: 52.76%  Conservative: 160
Best Local Simlarity: 34.37%  Mismatches: 352
Query Match: 27.94%      Indels: 59
DB:          1          Gaps: 22
US-10-030-330-1 (1-843) x POYAAAA (1-2661)
Qy      1 MetLysLysSerPheLeuLeuAlaIleValMetLeuPheGluIleAlaMetGlnLysHis 20
Db      55 ATGAACGATATCTTACACCTTAGGGCTTATTTACTGTGTCCTATGCTC---CAG 111
Qy      21 SerAlaProVal1ThrLysGluArgAlaLeuSerleuAlaArg-----leuAlaLeuArg 38
Db      112 GCAGGACCGGTGACAGATATCAACGCGAAGACAGCGCTAAGAACTTTTGGCCAAACGA 171
Qy      39 GlnValSerLeuArgMetGlyGlnThrAlaValSerAspLysLysIleSerIleAspTyrVal 58
Db      172 CAACCCACGTTG-----TCTTATCAACACCGAGTCTCCGGATGATTTGCTT 219
Qy      59 TyrArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGluGlySerProAlaTyr 78
Db      220 TACAAAGCTGCACAAAGAGAGAG-----GCACATA 249
Qy      79 PheTyrVal1AlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuValAlaAlaAspArg 98
Db      250 TTCTGCTGTTTCAATGAGAGAGAGAAAGACGATTTCTCTCTCGGAGGATGATCGG 309
Qy      99 IleProThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAsp 118
Db      310 TTCCCGAGGTGATCGGATATGCTTTCACAGGGGCACTTCATCGCGCCCTATGCGGAC 369
Qy      119 AsnLeuArgMetTrpLeuGlnIleTyrAspGlnGluIleGlyLeuLeuLeuSerGlyLys 138
Db      370 AATCTCAGGGGCGTGCCTCAAGGCTATGAACGTGAATGCTGTCTGTAATGAGCGGCAAG 429
Qy      139 AlaGlnLeuAsnGluLysIleuArg-----ThrGluGlyValProAlaGluVal 155
Db      430 GCAGACCGCATGATATCTTACCTCGTGAAGCCACCACTACACGGGACCTGCATCATCAT 489
Qy      156 HisAlaLeuMetAspAsnGlnLysPheAlaAsnAspPheMetArgTrpAsnGlnLys 175
Db      490 GCCCTTATTTGGAAACGGGCGAACAATGCAATCGATCGATCTTGTGGATCAGGGCTAT 549
Qy      176 ProTrpAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysVal 195
Db      550 CCATTTAACACCTGTGATCCCTGCTTCCTCCGGGAGAGACGCTTATACGGTGTGTT 609
Qy      196 AlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTrpProLeuGlnGlyGlySer 215
Db      610 GCCACCGCGCATGAGCAATATCATGCTACATCAAGTGGCGGAGGAGCTTCGGCGAA 669
Qy      216 PheAspTyrHisAlaGlySerleuValGly-----AsnTrpSerGlyThrPheGly 232
Db      670 TATGACTACTAC---GATGATATGACGGGAGACCAATACCACTATTCAGGCATTCGGT 726
Qy      233 GluMetTyrAspTrpIleAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSer 252
Db      727 GATACCTATTAACATGAGCAAGATGCCGGTAAATCATGATGTT---GATATCACCCCTGAA 783

```

OY	253	GlnValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerValSermetsSerPhe	272
Dd	784	GAAGTAAGAAGCCTTAGACTTCATGAGATGTGCACACTTTCTTGCAATAATGCAGATT	843
OY	273	TyrGluaSnGlySerGlyThrTyrSerValItyrValValGlyAlaLeuArgAsnAspHe	292
Dd	844	GCCGATTTTCGGCACAGTGGCATTTTCTCCCAATATTGTGCGAAGCGCCCTGGGAGACCTTT	903
OY	293	ArgTyrLysArgSerLeuGluLeuHisValArgAlaLeuTyrThrsSerGluITPrHis	312
Dd	904	CATTCACAAGAAATGCCSTTCGTACATCCATGCGCTCTCTCCCGGTAAAGATGGAAA	963
OY	313	AspMetIleArgGlyGluLeuAlaSerGlyArgProValITyrrAlaGlyAsnAsnGln	332
Dd	964	GACATGATCTTCAAGCACTGGCAGAAAAACAGACCTCGTATTATGCGCGGTGCAGACGGA	1023
OY	333	SerIleGlyHisAlaPheValCysAspGlyItyrAlaSerAspGlyThrsPheHisPheAsn	352
Dd	1024	TCGATGGGACATGCTTTGTTGGCANTGGATACGAAACAGACGGAACCTTCACCTCAAC	1083
OY	353	TyrGlyTrpGlyGlyValSerAsnGlyPheTyrLysLeuThrLeuSerProthrSer	372
Dd	1084	TGGGATGGGGTGGCATAGTAAAGCGTAATTTCTATCTCATCTCACTCAATCCGGGTTCG	1143
OY	373	LeuGlyIleGlyGlyGlyIleGlyPheThrIleTyrGlnGluIleIleThrGlyIle	392
Dd	1144	CTTGGCACAGCGCGGAGGTAGAGGCTACTCTAGCACAAAGAGTTGTGCATAGGCATA	1203
OY	393	GluProAlaItyrThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLys	412
Dd	1204	GAGCGGCGCAGCATGAAGCA---CCGGATTCGTACCAGATCCGACCATCAGCTCAT	1260
OY	413	AspIleGluAlaGluTyrLysSerGluSerGlyLeuAsnValGlyTyrSerIleTyrAsn	432
Dd	1261	GGATTGGCAA---CAACAATATGTCTGATAGACACTGACCTGAGCGTAAATCAAGAAC	1317
OY	433	ThrGlyGlnGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlu	452
Dd	1318	TACTCCACTTATWCGAGGAGATGTGAATATGGCTTATCCGCTGACGCTTCCCAATGAACG	1377
OY	453	ValIleGlyValLysThrSerSerIleAsnIleSerTPryGly---TygGlyGlnHis	471
Dd	1378	GAAACGACTAACCCGCGCTACCTCCCTATCGTTGGGAGACATTATCGAGAAAGC	1437
OY	472	ProGluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeu	491
Dd	1438	ACGGGCAACATTAACATACATACATGCAGCTAATTTGCCGAGGGGAAAAAATCCATCTCCATA	1499
OY	492	LeuTyrAlaArgTgtHglyThrGlnGlnTrpGluProValArgHisAlaGlnGlyItyr	511
Dd	1498	CTCTATTCACAAAGATGAGTGGCAGATTGGAAAGAGCTTAAGCAATCTATTAGGGGCTT	1557
OY	512	ValAsnSerIleLysValAsnThrThrAspProAsnAsnValValAlaThrValAspAsn	531
Dd	1558	GTCATTAAGATGAAGTGT---ACTATGCTCTCCGGAGTGAAGCTTACTCGGTT-----	1608
OY	532	AsnGlnGlyLysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyrGlu	551
Dd	1609	GCCGATGCCAGAGTGTCTTCACAAAGCGGTCTCTTTCGCACAMCCTAAGACATATTC	1668
OY	552	HisSerThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThrProval	571
Dd	1669	GACTGCAACACACAGTGCACAGTCTACAAACCCGGGTACGAAAGAAATTCGTAGTCGCGTC	1728
OY	572	AlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyITyrValMet	591
Dd	1729	ACCTTTGGCCCTCCGCAAT-----ACAAGGAGCGGCTCTATTCCTTGGGAGACATTTA	1782
OY	592	AlaGluVal---ProGlyItyrSerSerAsnTyrProValValITrpSerLysAspValLeu	610
Dd	1783	GTACAAATTACACCGGGGAGATGAAGACCGCGCAAAAAGTTTGGCTGACATAATACAGGCTTC	1842

Qy		611	ThlIleuSerGIuSglYAsprTYrThLleuTrpYrUg-----PhseRIeIAAsnGln	628
Dd		1843	AAGCGTCGTGCAGGCAAAATACATGGCTTCTGTACAGGGCCGAATGAAATTCTTAATGGAA	1902
Qy		629	LysAspGIurPlYsYIleGlySevAlserValIysThrProThGlutYrThnIs	648
Dd		1903	GATGCCCATGGATATGAGTTGGCTTCCATAGAATGAGAGAGCATACAGACCATCA	1962
Qy		649	ProLeuRheGluValGIynIsAAsnIThSeriThYrThrLeuAspMetAlnIsAn	668
Dd		1963	TGCTTACTGGTGCGCCTCCAACCCACAGATCATCTTCTCACAGTTGATCGGGCCAATCCC	2022
Qy		669	ArgValLeuProAspRhetHrLeuIlysAnIeu---GylLeuProRheAsnIGluIeu	687
Dd		2023	GAGACATTCGCGACTTTCAGCATTTACAATAAGAGAGGTGGCTGACTTCTCCGGAAAATTC	2082
Qy		688	ValValValPheArgGlnThrGlnSerSevSergIyseLeuTrpAlaIalGlnIuthr	707
Dd		2083	GAATATAGTGGCTATTAAGGCTTTCTCGGAAACT---TTCCTTCAAAGCCAAAGMAACAC	2139
Qy		708	ValnIsIleIysGInglYgluthrPheValTyrrIysProvalValGluGlyPro----	725
Dd		2140	ATGAGCTCTCGCCCAAGGGGAAACCAAAATATTGTCTCCGGAGCTGACTCGCAACTCTCT	2199
Qy		726	-----ILeropsrGIyseTyraGlyAlaThrlLeuHISAlaRhe	738
Dd		2200	CTCATATCAAAATGCCGAACCTCTTCCCATGGCATGTATACATTGTCAAGA-----	2253
Qy		739	ValAsnGIyGInglInIleuTYr-----LeuIysGIyIysArgAsnTYrThr	754
Dd		2254	-----GAGCAGAGGATTTTGGGATGCCATTCGATTTGTTGGGAGCTATTACTAGTCT	2304
Qy		755	ValIysIleValAsnGIYThrAlaValGluAlaIleGIuseRserGIu-----GluIle	772
Dd		2305	ATCCGCTCATATTACGAGTCTATCCCTCTCGGADATCGCTGGTAAGAGATGTTTCACTATA	2364
Qy		773	ArgValAPheProAspProAlaArgAspTyrrAlaGluIleSerAlarProCyIIleProGln	792
Dd		2365	GTACTTATTCCTCAATCTGCTGCACAGCTATGTCCATGTAGCATTCCTCCACATATGCC	2424
Qy		793	GIurThrSerIleIleLeuPheAspRheuSergIyIysIleValMetIysAsnSerLeu-Se	812
Dd		2425	GCGAGCACACTTCCTGTTTGTTCGATTTTCAAGGGCCAAATGCAAGCTCCACAGAAATCGCA	2484
Qy		812	rAlaGIynIsGlyArMeTasPvaIserArgLeuProAsnGluAlaTyrrIleLeuIysVa	832
Dd		2485	ATCTGCCATATAGCGCTCTCGACGTGCAGACGCTTCCGAAGGGCACACATATCGTGTGTG	2544
Qy		832	IAspGlyTyrrThrThryIysIleAsnIle 841	
Dd		2545	AGAAAGACATGTAGGGAAGCTCTTTATT 2572	
RESULT 3				
LOCUS	S75942	4056 bp	DNA linear	BCT 27-JUL-1995
DEFINITION	pTci-99 kDa cysteine protease/hemagglutinin [Porphyromonas			
ACCESSION	S75942			
VERSION	S75942.1	GI:913136		
KEYWORDS	.			
SOURCE	Porphyromonas gingivalis ATCC 53977.			
ORGANISM	Porphyromonas gingivalis Bacteria; Bacteroidetes; Bacteroides; Bacteroidales; Porphyromonadaceae; Porphyromonas.			
REFERENCE	Madden,T.E., Clark,Y.L. and Kiranltzu,H.K. Revised sequence of the Porphyromonas gingivalis prtI cysteine protease/hemagglutinin gene: homology with streptococcal pyrogenic exotoxin B/streptococcal proteinase infect. Immun. 63 (1), 238-247 (1995)			
JOURNAL MEDLINE	95105001			
PUBMED	7806362			
REMARK	Genbank staff at the National Library of Medicine created this			


```

Oy 472 ProgluSerPheSerLeuAlaProangInLeuSerGlnGlyIleAsnThrIleu 491
Db 2203 ACGGCAACATACATACATACATACATACATACATACATACATACATACAT 2262
Oy 492 LeuTyraGatGthGlyThrluIntrpGluProValArgHisAlaGlnGlyTy 511
Db 2263 CTCTATCGAAGACAGATGCGATGCGAAGGAAAGGAAAGGAAAGGAAAGG 2322
Oy 512 ValaSerIleValValaSerThrAspProAsnValValValValValAsp 531
Db 2333 GTCATATAGATAGAAAGT---ACTATGCTGCGGAGAGATGATGCTTACCG 2373
Oy 532 AangluGlyLeuLeuSerIleValProaSerPheValAlaAspLeuAsnSer 551
Db 2374 GCCGATGCCAGAGATGCTCTCAAGACGCTTCTTTCGACACATGAACATAT 2433
Oy 552 HisSerThrIleThrValGlnPheAsnSerAspSerProAspGluIleArg 571
Db 2434 GACTCCAAACCTAGTGGCCACAGCTACAAACCCGGGTACGGAAGAAATTCGT 2493
Oy 572 AlaPheAlaLeuSerThrGlyValaThrAlaAspValIleSerLeuGlyTrp 591
Db 2494 ACCTTGGCCCTCCGCAAT---ACAGAGGAGCGGCTCTATTCCTGGCAGACA 2547
Oy 592 AlaGluVal---ProGlyGlySerSerAsnTyrrProValValTrpSerLys 610
Db 2548 GTAGAAATTAACACCCGCGAGATAGACGCGGAAAGAAATTCGCTCAAT 2607
Oy 611 ThrLeuSerGluGlyAspTyrrThrLeuTyrrTyrrArg---PheSerIle 628
Db 2608 AAGCTCTGTCGACGACATACATGCTTGTCTACGCGCGATAGATGGAATTC 2667
Oy 629 LysAspGluTrpLysLysIleGlySerValSerValLysThrProThrGlu 648
Db 2668 GATGCGCTATGATAGATGATGCTTCCATAGAAAGACAGACATACAGACCA 2727
Oy 649 ProLeuPheGluValGlyHisAsnGlnThrSerThrTyrrThrLeuAsp 668
Db 2728 TCGTACTAGTGGCTCCCAACCCACAGATCGATCTCTCAGTATCGGCGCA 2787
Oy 669 ArgValLeuProAspPheThrLeuLysAsnLeu---GlyLeuProPheAs 687
Db 2788 GAGACATCCGACCTTTCAGCATTTACMAATGAGGTGGTCTTCTCCGGAAT 2847
Oy 688 ValValValPheArgIleThrGlnSerSerGlySerLeuTrpAlaAlaGln 707
Db 2848 GAATATAGTGGTATTAAGGCTTTCGGAACCT---TTCCTCAAGCGAAAGA 2904
Oy 708 ValHisIleLysGlnGlyGluThrPheValTyrrLysProValValGlu 725
Db 2905 AAGACTTCGCCCAAGGCGAAACAAAGTATGCTCTCGGACCTGACTCGA 2964
Oy 726 -----IleProAspGlySerTyrrArgAlaThrIleuHisAlaPhe 738
Db 2965 CTCTATACAAATCCGAACTTTCCGATGCGCATCTATTTACATTTGTCACA 3018
Oy 739 ValaAngluGlnGlnLeuTyrr-----LeuLysGlyLysArgAsnTyrr 754
Db 3019 -----GACGAGGATTTGGGATCGCATGATTTGTTGGGACATATACAT 3069
Oy 755 ValLysIleValaAsnGlyThrAlaValGluAlaIleGluSerGlu----- 772
Db 3070 ATCCGCTCATACGAGATCTATCTCTTCGACATCGCTGTAAGAGATTTCT 3129
Oy 773 ArgValPheProAspProAlaArgAspTyrrValGluIleSerAlaProCys 792
Db 3130 GTACTTTATCCCAATCTGCTACAGACTATGTCATGTAAGCATTCCTCCACA 3189
Oy 793 GluThrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLysAsn 812
Db 3190 GGCACACACACTTCTGTTGTCATATTCAGAGCGCAATTCACCTCCACGAAG 3249
Oy 812 fAlaGlyHisGlyArgMetAspValSerArgLeuProAsnGluAlaTyrrIle 832

```

```

Db 3250 ATCTCCCATATGCTCTTCGACAGTCAAGCTTCTCCGAAGGCGACCTAT 3309
Oy 832 lAspGlyTyrrThrThrLysIleAsnIle 841
Db 3310 AGAAGACATGTAAGGAAAGCTCTTATAT 3337

RESULT 4
STRPEBJ
LOCUS
DEFINITION
ACCESSION L26134
VERSION L26134.1
KEYWORDS GI:431655
SOURCE
ORGANISM
Streptococcus pyogenes.
Streptococcus pyogenes.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 1357)
REFERENCE
1 Kapur V., Topouzis S., Majesky M.W., Li L.L., Hamrick M.R.,
Hamill R.J., Patel J.M. and Musser J.M.
A conserved Streptococcus pyogenes extracellular cysteine protease
cleaves human fibronectin and degrades vitronectin
Microb. Pathog. 15 (5), 327-346 (1993)
94285789
FEATURES
source
location/Qualifiers
1..1357
/organism="Streptococcus pyogenes"
/isolate="NGAS 587"
/db_xref="taxon:1314"
161..1357
/gene="speB"
161..1357
/gene="speB"
/function="cysteine protease precursor"
/note="allele 18"
/codon_start=1
/transl_table=11
/product="pyrogenic exotoxin B"
/protein_id="AA027000.1"
/db_xref="GI:431656"
/translacion="MNRKLLGRLSLALNGFTVLANFVADQNFARNEKEKKSAT
FIQKSAIKAGARSADIKLDKVNIGGLSGSNMIVYI STGEFIVSGDRSPILG
YSTSGSPANGKENIASFMESEYEOIKENKIDTACTAEIKQIVVSLDSKGIHY
NOGNFYNLTPYIEKVKPGEOSEFVGOHAKTCGVATPAQIMKHYHYPKRGKDYTL
SSNPFYHNPRLFPALISTROVMNNIILPTSGRSNVOKMAISLMDVGISVMDY
GSSGSASSRVRQRLRENFQGNVSHINRSDPKOMESQIDKELSONOPRYQGV
GVGSHARVIDGADGRNFYHVMGWSVSDGFRLDALNPSALGTGGAGGNGTOSA
VVGIRP"
BASE COUNT 446 a 262 c 267 g 382 t
ORIGIN
Alignment Scores:
Pred. No.: 3.3e-24 Length: 1357
Score: 499.00 Matches: 126
Percent Similarity: 50.13% Conservative: 72
Best Local Similarity: 31.90% Mismatches: 139
Query Match: 11.24% Indels: 59
Gaps: 12
US-10-030-330-1 (1-843) x STRPEBJ (1-1357)
Oy 26 LysGluArgAlaLeuSerLeuAlaArgValaLeuArgGlnValSerLeuArg 45
Db 267 AAGAAACAAAGATAGCGCTATCATCAATTAACAAATACACA-GCTATCAAGCAG 325
Oy 46 GluThrAlaValSerAspLysIleSerIleAspTyrrValTyrrArgGlnGly 65
Db 326 GCACGACGCGCAAGAT---ATTAAAGCTTGCAAGAAAGTTAAGTGGAGAACTT 379
Oy 66 ArgGlyIleThrSerGlnGlnGlySerProAlaTyrrPheTyrrValaIleAsn 85

```

```

Db 380 -----TCTGCTCTAATATCTATCTTACAAATTTCTACTGCA 418
Qy 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspArgIleProThrIleLeuAlaTyr 105
Db 419 -----GGATTGTTATCGTTTCAGGAGATAAACGTTCCCAAAATTTAGAGATAC 469
Qy 106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGln 125
Db 470 TCTACACGCGATTCATTGACGCTAACGCT---AAGAAACCACTTCCTCTTCACAGGA 526
Qy 126 ILeTyr-----AspGlnIleLeuIleLeuSerGlyTyrAla 139
Db 527 ACTTATGTCGAACAATCAAGAAATAATAATAGACATCTATCTGCTGACCCT 586
Qy 140 GlnLeuAsnGluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMet 159
Db 587 GAGATTAAACCAACCGTT-----GTAAATCTCTCTCT 619
Qy 160 AspAsnGlyHisPheAlaAsnAspProMetArgTrpAsnGlnGlyTyrProTrpAsnAsn 179
Db 620 GAT-----TCAAAAGCATTCATTACAAATCAAGTAACCTTACAACTGA 664
Qy 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
Db 665 TTGACACCTGTTATTGAAAGTAACAAACAGGTGAACATCTTTGTAGTCAAGTGA 724
Qy 191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTrpLeu 210
Db 725 GCTACAGATGTGTGCTACTGCACTGCTCAAAATATGCAATATCATATATCCCTAAC 784
Qy 211 GlnGlyGluGlySerPheAspTyr-----HisAla 220
Db 785 AAGGCTTGAAGACTACTACTTACACATCAAGTCAAAATATTCATTCACATCCCT 844
Qy 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsn--- 239
Db 845 AAGAACTTG-----TTTGACAGTATCTGCTACTAGACAAATACAACTGGAACAACTC 895
Qy 240 MetTrpGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
Db 896 CTACTACTTAAAGCGAAGAGAAATCTAACTGTTCAAAAAATG-----GCGATTTCAGAA 949
Qy 260 LeuMetArgAspValSerAlaSerAlaSerMetSerPheTyrGluAsnGlySerGlyThr 279
Db 950 TTGATGCGCTGATGTTGCTATTCAGTACATGATGATTAATGCTCATCTAGTGTCTGCA 1009
Qy 280 TyrSerValTyrValGlyAlaLeuArgAsnAsnAspPheArgTyrIleLysArgSerLeuGln 299
Db 1010 GGTACGCTCTGTTCAAGACGCTGAGAGAAACTGTTGGCTACAAACAACTGTTTCA 1069
Qy 300 LeuHisValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyLeu 319
Db 1070 CAATTCACCCGATGACTTGTAGCAAAAGATGGGAATTCACAAATTTACAAAGATTA 1129
Qy 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
Db 1130 TCTCAAAACCAACGATATCTACCAAGTGTGCGTAAGTAGCGGACATGCTTGTGT 1189
Qy 340 CysAspGlyTyrAlaSerAspLysIleSerIlePheHisPheAsnTrpGlyTyrPheGlyValSer 359
Db 1190 ATCGATGCTGTGACGAGCGATCACTTACATGTTAACTGGGGTGGAGTGTCT 1249
Qy 360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGly 379
Db 1250 GACGGCTCTCTCGTTGACGACATAAACCTTGTGCTGTGTAAGTGTGGCGGCA 1309
Qy 380 IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394
Db 1310 GCGCGCTTACAGGTTTACCAAAAGTGTGTGTAGGATCAAACT 1354

```

RESULT 5
STRSPEBO

```

LOCUS STRSPEBO 1357 bp DNA linear BCT 18-MAY-2001
DEFINITION Streptococcus pyogenes pyrogenic exotoxin B (speb) gene, complete
ACCESSION L26139
VERSION L26139.1 GI:431665
KEYWORDS
SOURCE
ORGANISM Streptococcus pyogenes.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus pyogenes.
REFERENCE
AUTHORS Kapur,V., Topolazis,S., Majesky,M.W., Li,L.L., Hamrick,M.R.,
Hamill,R.J., Pettit,J.M. and Musser,J.M.
TITLE A conserved Streptococcus pyogenes extracellular cysteine protease
cleaves human fibronectin and degrades vitronectin
JOURNAL Microb. Pathog. 15 (5), 327-346 (1993)
MEDLINE 94285789
PUBMED 7516997
FEATURES
source location/Qualifiers
1..1357
/organism="Streptococcus pyogenes"
/isolate="MGAS 650"
/db_xref="taxon:1314"
161..1357
/gene="speb"
161..1357
/gene="speb"
/function="cysteine protease precursor"
/note="allele 23"
/codon_start=1
/tranal_table=11
/product="pyrogenic exotoxin B"
/protein_id="MAN27005.1"
/db_xref="GI:431666"
/translaction="NMNRKRLGRLSLALGCGVLANPVFADONFARNEKANDSAIT
FIKQSAIRAGARSADIEDIKVNLGEGESGSMYYNISTGFIYSGDKRPEILG
YTSQSGFDANGKENIASPDESYEOKENKKIDPTIYGAIEIKOPVFKSLDKGTHY
NOGNPNYLLTPVIEKYKPGEOSVGOHAAATATATDQIMKYHNPNGGIDQYTYL
SSNPNFNNPKNLFALISTROYMNNILPTYSRESNVOKMAISELMADVGISYVDY
GFSGSGASSRYORALKENGTINOSYHOIHRSDFSKQDEADQIDKELSDQPYTQYG
GAVGGAHAFVIDGADGRNFYHVMWGVSDGFRLDALNPALGTTGGAGGNGYQSA
VVGIRK"
BASE COUNT 446 a 264 c 268 g 379 t
ORIGIN
Alignment Scores:
Pred. No.: 3..3e-24 Length: 1357
Score: 499.00 Matches: 126
Percent Similarity: 50.388 Conservative: 73
Best Local Similarity: 31.908 Mismatches: 138
Query Match: 1 Indels: 59
DB: 1 Gaps: 12
US-10-030-330-1 (1-843) x STRSPEBO (1-1357)
Qy 26 LysGluArgAlaLeuSerIleuAlaArgLeuAlaLeuArgGlnValSerLeuArgMetGly 45
Db 267 AAGACCAAAAGATACCGCTATATCATTTATCCAAAATACAGA-GCTATCAAGCAGGT 325
Qy 46 GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu 65
Db 326 GCACGAAGCGCAGCAAGAT---ATTAAAGCTTGAACAAAGTTAACTTAGGTGAGAACT--- 379
Qy 66 ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
Db 380 -----TCTGCTCTAATATCTATCTTACAAATTTCTACTGCA 418
Qy 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspArgIleProThrIleLeuAlaTyr 105
Db 419 -----GGATTGTTATCGTTTCAGGAGATAAACGTTCCCAAAATTTAGAGATAC 469
Qy 106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGln 125

```

```

Db 470 TCTACACGGGATCATTTGACGCTAACGGT---AAGAAAAACATTCCTTCTCATCGAA 526
Oy 126 lIeTYr-----AspGInGluIleGlyLeuIleuSerGlyLysAla 139
Db 527 AGTTATGTGGAACAAATCAAGAAAAATTAAGACATTAATTCCTGCTACCGCT 586
Oy 140 GInLeuAsnGluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMet 159
Db 587 GAGATTAAACACACAGTT-----GTTAAATCTCTCCCTT 619
Oy 160 AspaAsnGlyHisPheHisAsnAspPromeLArgThrPasnGInGlyTYrProThrPasnAsn 179
Db 620 GAT-----TCAAAAGCATTCATTAACAATCAAGTAACCCCTTACCAACCTTA 664
Oy 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
Db 665 TTGACACACCTGTTATTTGAAAAAGTAAACAGGTGAACAATCTTTTGTAGCTCAACCTGCA 724
Oy 191 TYrThrGlyCysValAlaThrAlaAlaAlaGInIleMetArgTYrHisSerTrpProLeu 210
Db 725 GCTACAGGATGTGTGCTACTGCTCAACTGCTCAAAATTAATGAAATATCATTAATACCTAAC 784
Oy 211 GInGlyGluGlySerPheAspTYr-----HisAla 220
Db 785 AAGGGGTGGAAGACTACACTTACACACTACCTCAAAATACCAATATTTCAACCATCCT 844
Oy 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTYrAspTrpLeuAsn--- 239
Db 845 AAGAACTTG-----TTGACAGCTATCTCTACTAGCAATACATCAAGTGAACATC 895
Oy 240 MetProGlyAsnProAspLeuAspAsnLeuThrGInSerGInValAspAlaTYrAlaThr 259
Db 896 CTACCTACTTATAGCGGAAGAGATCTAACGTTCAAAAAAAGT-----GCGATTTCAGAA 949
Oy 260 LeuMetArgAspValSerAlaSerValSerMetSerPheTYrGluAsnGlySerGlyThr 279
Db 950 TTGATGGCTGATGTGTGATTTTCAGTACATGATTAATGATTCACATGATGCTTCTGCA 1009
Oy 280 TYrSerValTYrValAlaGlyAlaLeuArgAsnAsnPheArgTYrLysArgSerLeuGIn 299
Db 1010 GGTAGCTCTCGTTCGTTCAAGACCTTGAAGAATACTTGGCTACACCAATCTGTTCAC 1069
Oy 300 LeuHisValArgAlaLeuTYrThrSerGInGluTYrPheHisAspMetIleArgGlyLeu 319
Db 1070 CAATATCAACCGTAGCGACTTAGCAAAACAAAGATTGGGAAGCAAAATTGACAAAGATTTG 1129
Oy 320 AlaSerGlyArgProValTYrTYrAlaGlyAsnAsnGInSerIleGlyHisAlaPheVal 339
Db 1130 TCTCAAAACCAACAGTATACATACCAAGGTGCGTAAGTAAGCGGACATGCTTTGTT 1189
Oy 340 CysAspGlyTYrAlaSerAspGlyThrPheHisPheAsnTrpGlyTYrGlyValAlaSer 359
Db 1199 ATCGATGGTGGTGCAGCGAGCTAACTTCTACATGTTAACTGGGGTGGAGTCTCT 1249
Oy 360 AsnGlyPheTYrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlySerGlyGly 379
Db 1250 GACGCTCTCTCGTTCGTTGACGACATAAACCTTCAGCTGTGCTGCTGCTGCGCGCA 1309
Oy 380 lIeGlyPheThrIleTYrGInGluIleIleThrGlyIleGluPro 394
Db 1310 GCGGCTTCAACGCTTACCAAGTGTGTTGATGATCAAACT 1354

```

RESULT 6
 STRSPEBO 1357 bp DNA linear BCT 18-MAY-2001
 LOCUS Streptococcus pyogenes pyrogenic exotoxin B (speB) gene, complete cds.
 DEFINITION
 ACCESSION L26141.1 GI:431669
 VERSION L26141
 KEYWORDS
 SOURCE Streptococcus pyogenes.
 ORGANISM Streptococcus pyogenes
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

```

REFERENCE 1 (bases 1 to 1357)
AUTHORS Kapur V., Topouzis S., Majesky M.W., Li L.L., Hamrick M.R., Hamill R.J., Patti J.M. and Musser J.M.
TITLE A conserved Streptococcus pyogenes extracellular cysteine protease cleaves human fibronectin and degrades vitronectin
JOURNAL Microb. Pathog. 15 (5), 327-346 (1993)
MEDLINE 94285789
PUBMED 7516997
FEATURES
    source
        1..1357
            /organism="Streptococcus pyogenes"
            /isolate="MGAS 168"
            /db_xref="taxon:1314"
            161..1357
            /gene="speB"
            161..1357
            /gene="speB"
            /function="cysteine protease precursor"
            /note="allele 25"
            /codon_start=1
            /transl_table=11
            /product="pyrogenic exotoxin B"
            /protein_id="AAZ7007.1"
            /db_xref="GI:431670"
    CDS
        161..1357
        /transl_table=11
        /product="pyrogenic exotoxin B"
        /protein_id="AAZ7007.1"
        /db_xref="GI:431670"
        /translation="MNKKRIGIRLSLALGFFVLANPVADQNFARNEKAKDSAIT
        FLOKSAATKAGARSADIEDIKDVNIGGELSGSNMTIYNISGTFVYISGDRSPETIG
        YSTGSPVYNGKENTASPPESVIEDIKKKIDPTIYAGTAELKQPVVSLDSKGIHY
        NQGNFYNLPIEYKVEQDSFVQGHATGCVATATQIMKTHNPNGKLDVITYLL
        SSNNPYFNHPKMLFAAISTROYNNMNIILPTSGRSNNVQKAIISLMDVGLISVMDY
        GSSGSGASSRIVORALKENFGVNOGHOIRNSDFKQDMSOIDELELSONGVYVQV
        GVGGHAFVIDADGRNFYHVMGCVSDGPFRLDALNPALGTGCGAGGNGYQSA
        VYGIKP"
BASE COUNT 448 a 265 c 265 g 379 t
ORIGIN
Alignment Scores:
Pred. No.: 3..3e-24 Length: 1357
Score: 499.00 Matches: 125
Percent Similarity: 50.38 Conservative: 74
Best Local Similarity: 31.65 Mismatches: 138
Query Match: 11.248 Indels: 59
DB: 1 Gaps: 12
US-10-030-330-1 (1-843) x STRSPEBO (1-1357)
Oy 26 LysGluArgAlaLeuSerIleuAlaGlyLeuAlaLeuArgGInValSerIleuArgMetGly 45
Db 267 AAGAAAGCAAAAGATACCGCTATCACATTTATCCAAAATCAACA-GCTATCAAGCAGGT 325
Oy 46 GInThrAlaValSerAspLysIleSerIleAspTYrValTYrArgGInGlyAspAlaGlu 65
Db 326 GCACGAGCGCAGCAAGAT---ATTAAAGCTTGACAAAGTTAACTTAGTGAGAAACT--- 379
Oy 66 ArgGlyIleThrSerGInGluGlySerProAlaTYrPheTYrValAlaAsnArgGly 85
Db 380 -----TCTGGCTCTAATATGATATATATTAACAATATTCTACTGGA 418
Oy 86 AsnAsnGluGlyTYrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTYr 105
Db 419 -----GGAATTTGTTATCGTTTCAGAGATAAACGTTCCAGAAATTTGTAGGATAC 469
Oy 106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGIn 125
Db 470 TCTACACGGGATCATTTTACGTTAACGTT---AAGAAATAATTCCTTCTCATCGAA 526
Oy 126 lIeTYr-----AspGInGluIleGlyLeuIleuSerGlyLysAla 139
Db 527 AGTTATGTGGAACAAATCAAGAAAAATTAAGACATTAATTCCTGCTACCGCT 586
Oy 140 GInLeuAsnGluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMet 159
Db 587 GAGATTAAACACACAGTT-----GTTAAATCTCTCCCTT 619

```


Oy 191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTrpProLeu 210
 Db |||||
 Oy 725 GCRACAGAGATGTCTTCTACTGCAACTGCTCAAAATTATGATATCATTAATTAACCTAAC 784
 Db |||||
 Oy 211 GlnGlyGluGlySerPheAspTyr-----HisAla 220
 Db |||||
 Oy 785 AAGGGCTGAAGACTACACTTACACACTTAAGCTCAATTAACCATTAATTCACACACT 844
 Db |||||
 Oy 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsn--- 239
 Db |||||
 Oy 845 AAGAACTT-----TTTGACGCTATCTCTACTAGACATTAACCAATGGAACACATC 895
 Db |||||
 Oy 240 MetTrpGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
 Db |||||
 Oy 896 CTACCTACTTATAGCGGAGAGATCTAACCTTCAAAAATG-----GCGATTTCAGAA 949
 Db |||||
 Oy 260 LeuMetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThr 279
 Db |||||
 Oy 950 TTGATGGCTGATGTGGTATTTCAAGTACATGATGATGATGATGATGATGATGATGATGAT 1009
 Db |||||
 Oy 280 TyrSerValTyrValValGlyAlaLeuArgAsnAspPheArgTyrLysArgSerLeuGln 1069
 Db |||||
 Oy 1010 GCRAGCTCTGCTGTCAAGAGCCTTGAAGAAAGAACTTGGCTACAAACCAATCTGTCTAC 1069
 Db |||||
 Oy 300 LeuHisValArgAlaLeuThrTyrThrSerGlnGluTyrPheAspMetIleArgGlyLeu 1129
 Db |||||
 Oy 1070 CAATCAACCGCTAGCGACTTTCAGCAACAAAGTGGAGATCAAAATTCAGCAAAAGATTA 1129
 Db |||||
 Oy 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 1139
 Db |||||
 Oy 1130 TCTCAAAACCAACAGTATCTACTACCAAGTGTGCTGTAAGAGCGGACATGCTTGT 1189
 Db |||||
 Oy 340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrPheGlyValSer 1190
 Db |||||
 Oy 1190 ATCGATGTGTGTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1249
 Db |||||
 Oy 360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGluGly 379
 Db |||||
 Oy 1250 GACGCTCTTCCCTCTTTCAGCAGCTAACCCCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309
 Db |||||
 Oy 380 IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGlyPro 394
 Db |||||
 Oy 1310 GCGGCTTCAACGCTTACCAAAAGTGTGCTGTAAGCAGCAACCT 1354
 Db |||||
 RESULT 8
 AB051298
 LOCUS
 ANNOTATION
 Streptococcus pyogenes spe b gene for pyrogenic exotoxin B,
 complete cds.
 AB051298
 AB051298.1 GI:13591590
 KEYWORDS
 Streptococcus pyogenes (strain: NZ131) DNA.
 SOURCE
 Streptococcus pyogenes
 ORGANISM
 Streptococcus pyogenes (strain: NZ131) DNA.
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE
 1 (sites)
 Watanabe, Y. and Ohkuni, H.
 Cloning and expression analysis of Streptococcal pyrogenic exotoxin
 B (Spe-B)/streptococcal cysteine protease (SCP)
 2 (bases 1 to 1479)
 Watanabe, Y. and Ohkuni, H.
 Direct Submision
 Submitted (16-NOV-2000) Yukino Watanabe, Nippon Medical School
 Institute of Gerontology, Department of Immunology and Infectious
 Diseases, Nakahara-ku Kosugi-1-396, Kawasaki-shi, Kanagawa
 211-8533, Japan (E-mail: wtnbykn-nms@umln.ac.jp,
 Tel:81-44-733-1821(ex.892), Fax:81-44-733-1877)
 FEATURES
 source
 1..1479
 location/Qualifiers
 /organism="Streptococcus pyogenes"
 /strain="NZ131"

gene
 CDS
 /db_xref="taxon:1314"
 154..1350
 /gene="spe b"
 154..1350
 /gene="spe b"
 /codon_start=1
 /transl_table=11
 /product="pyrogenic exotoxin B"
 /protein_id="BAB40954.1"
 /db_xref="GI:13591591"
 /translation="MKKKLGRLLSLALGFLVLANPVPADONFANEREADDSAIT
 YKSSGFDVNGKRSKEDIKLVNIGELSGSNMXYVNISSGVYIVSGKRSPELIG
 NGRFPLNLTPEVIEKVKPEGSFVEQIKENKLLDTTQAEIKQPVKSLDSKGTH
 SSNNPFYNNKPLFAIISTROYNNNNILPYSGRESVQKMAISELMADVGISVDKY
 GSSSSAGSSRYORALKENFGYNGSVYINRSDSKODMESQIDKELSONOPYOGV
 GVGGHAFVIGADGRNFRHYHNMGMGVSDGFRLDALNPSALGTGCGAGFNQVQSA
 VVGIRP"
 BASE COUNT
 490 a 298 c 288 g 403 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4 25e-24 Length: 1479
 Score: 498.00 Matches: 127
 Percent Similarity: 50.50% Conservative: 74
 Best Local Similarity: 31.91% Mismatches: 138
 Query Match: 11.22% Indels: 60
 DB: 1 Gaps: 13
 US-10-030-330-1 (1-843) x AB051298 (1-1479)
 Oy 24 ValThrLysGluArg---AlaLeuSerLeuAlaArgLeuAlaLeuArgGlnValSerLeu 42
 Db |||||
 Oy 251 GTACAGAAAAGAGGACAGCAAGATAGCGCTATCACATTATTCAGAAAATCAGCA-GCATC 309
 Db |||||
 Oy 43 ArgMetGlyGlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGly 62
 Db |||||
 Oy 310 AAAAGCGGTGACGAGGACGAGAGAT---ATTAAAGCTTCAACAAGTAACTTAGGTGA 366
 Db |||||
 Oy 63 AspAlaGluArgGlyIleThrSerGlnGluGlySerProAlaTyrPheThrValAla 82
 Db |||||
 Oy 367 GAACCTT-----TCTGGCTCTAATATGATGATTTTACATATT 402
 Db |||||
 Oy 83 AsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgLeuProThrIle 102
 Db |||||
 Oy 403 TCTACTGGA-----GGATTGTATGCTTTCAGAGATTAAGCTTCTCCAGAAATY 453
 Db |||||
 Oy 103 LeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMet 122
 Db |||||
 Oy 454 CTAGATACCTCTACAGCGGATCATTTGACGTTAACCGT---AAAGAAACATTGCTTCC 510
 Db |||||
 Oy 123 TrpLeuGlnIleTyr-----AspGlnGluIleGlyLeuIleLeuSer 136
 Db |||||
 Oy 511 TTCATGAAAGATTATGTCGACAAATCAAGAAACAAACAAATTAAGCACTTATGCT 570
 Db |||||
 Oy 137 GlyLysAlaGlnLeuAsnGlnGluIleLeuArgThrGlnGlyValProAlaGluValHis 156
 Db |||||
 Oy 571 GGTACCGCGAGATTAACCAACGATT-----GTTAA 603
 Db |||||
 Oy 157 AlaLeuMetAspAsnGlyHisPheAlaAsnAspPheMetArgTrpAsnGlnGlyTrpPro 176
 Db |||||
 Oy 604 TCTCTCTGTGAT-----TCAAAAGCATTCATTACATCAAGTAACTT 648
 Db |||||
 Oy 177 TrpAsnAsnLysGluProLeuLeuProAsn-----Gly 187
 Db |||||
 Oy 649 TACAACTTATGACACCTGTTATGAAAAGTAAACCGAGTAACTTTTGTAGCT 708
 Db |||||
 Oy 188 AsnHisAlaTyrThrGlyCysValAlaThrAlaAlaAlaGlnIleMetArgTyrHisSer 207
 Db |||||
 Oy 709 CAACATCAGCTACAGAGATGTCTTCTACTGCAACTGCTCAATATGAAATATCATATAT 768
 Oy 208 TrpProLeuGlnGlyGlnGlySerPheAspTyr----- 218

Db 769 TACCTAACAAAGGTTGAAAGCTACCTTACACACCTAAGCTAAATACCAATCTTTC 828
 QY 219 ---HISAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetGlyAsnTrp 237
 Db 829 AACCAATCTTAAGAACTG-----TTTGACAGTATCTCTACAGAACAAACACTGG 879
 QY 238 ILeaen---MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAla 256
 Db 880 AACAAACATCTCTACCTATATAGCGAAGAAATCTAACGCTCAAAAATG-----GCG 933
 QY 257 TTTAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheGlyGlnGly 276
 Db 934 ATTTCAGATTCATGCTGATGCTGATATCTAGTACAGATGATGATGATGATCTACT 993
 QY 277 SerGlyThrLeuSerValGlyValGlyAlaLeuAsnAsnPheArgGlyTrpAsnArg 296
 Db 994 GCTTTCGACGAGTACCTCTCTGCTTCAAAAGCCTTGAAGAAACCTTGGCTTCAACAA 1053
 QY 297 SerLeuGlnLeuValArgAlaLeuValLeuValLeuValLeuValLeuValLeuVal 316
 Db 1054 TCTGTTCACAAATACACGCTTACCTTACCAAAAGATGGGAATCACAAATTCAC 1113
 QY 317 GlyGluLeuAlaSerGlyArgProValGlyValGlyAlaGlyAsnAsnGlnSerIleGlyHis 336
 Db 1114 AAGAAATATCTCAAAACCAACCACTATCTACCAAGGTGCTGCTGCTGCTGCTGCTGCT 1173
 QY 337 AlaPheValCysAspGlyThrValAlaSerAspGlyThrPheHisPheAsnTrpGlyTrpGly 356
 Db 1174 GCTTTGTATGATGAGTGGTGGTACGACGACGACGACGACGACGACGACGACGACGACG 1233
 QY 357 GlyValSerAsnGlyPheThrIleValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 376
 Db 1234 GAGAGCTCTGACAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1293
 QY 377 GlyGluGlyIleGlyPheThrIleValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 394
 Db 1294 GCGCGCGAGCGGCTTCAACGCTTACCAAAAGTCTGCTGCTGCTGCTGCTGCTGCT 1347
 RESULT 9
 STRSPBAD 1357 bp DNA linear BCT 18-MAY-2001
 LOCUS Streptococcus pyogenes pyrogenic exotoxin B (speb) gene, complete cds.
 DEFINITION L26154
 ACCESSION L26154.1 GI:431621
 VERSION 1
 KEYWORDS Streptococcus pyogenes.
 SOURCE Streptococcus pyogenes.
 ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
 REFERENCE 1 (bases 1 to 1357)
 Kapur,V., Topouzis,S., Majesky,M.W., Li,L.L., Hamrick,M.R., Hamill,R.D., Patti,D.M. and Musser,J.M.
 A conserved Streptococcus pyogenes extracellular cysteine protease cleaves human fibronectin and degrades vitronectin
 JOURNAL Microb. Pathog. 15 (5), 327-346 (1993)
 MEDLINE 94285789
 PUBMED 7516997
 FEATURES
 source
 location/Qualifiers
 1..1357
 /organism="Streptococcus pyogenes"
 /isolate="MGAS 1882"
 /db_xref="taxon:1314"
 161..1357
 /gene="speB"
 161..1357
 /gene="speB"
 /function="cysteine protease precursor"
 /note="allele 37"
 /codon_start=1
 /transl_table=11
 /product="pyrogenic exotoxin B"
 /protein_id="AAA26983.1"

/db_xref="GI:431622"
 /translation="MNKKLGIIRLISLALGGFVLNPFVADONFARNEXEAKDSALT
 FTKSAATKAGARSADIKLIDVNIAGELSGSNMYYNIISTGGFPIVSDKSPELTG
 YSTSSFDVNGKENLASEMESTVEQIKENKLDITTYAGTAELKOPVNSLDSKGIH
 NOGNPNLPTPIEVRKPEQSFVQGHAAATGVAATAQIMKHNPNGGLDYITL
 SNNPNFNLKMLFAAISTROYNNNNILPTVGGRESNOVKAAISELMADVGISVDY
 GPSSGASRVRQALKEENFGYNOSVHQRNDFSDQDEADQIDELSONQVYQGV
 GKVGHFAPFIDADGRNPFYHVMGNGVSDFFRLDALNPSALGTGGCAGGRNGVQSA
 VYGIKP"
 BASE COUNT 448 a 263 c 266 g 380 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5, 21e-24
 Score: 496.00
 Percent Similarity: 50.388
 Best Local Similarity: 31.654
 Query Match: 11.184
 DB: 1
 Gaps: 12
 US-10-030-330-1 (1-843) x STRSPBAD (1-1357)
 QY 26 LysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuValGlnValSerLeuArgMetGly 45
 Db 267 AAGAAACAAAGATAGCGCTATCATTCATCAAAATCAGCA--GCTATCAAAAGAGGT 325
 QY 46 GlnThrAlaValSerAspIleSerIleAspTrpValValArgGlnGlnAspAlaGly 65
 Db 326 GCACGAAGGCGCAAGAAAT--ATTAGCTTGCACAAAGTAACTTAGCTGAGAACTT-- 379
 QY 66 ArgGlyIleThrSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 85
 Db 380 -----TCTGCTCTAATATATATATCTTACAAATATTTCTACTGGA 418
 QY 86 AsnAsnGlnGlyThrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTrp 105
 Db 419 -----GAAATTTGTATCTCTTTCAGAGAAATTAACGTTCCAGAAATTCAGAGATAC 469
 QY 106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGln 125
 Db 470 TCTACACGCGGATCATTTACGTTAACGCT--AAAGAAACATCTGCTTCTCATAGGAA 526
 QY 126 IleTyr-----AspGlnGlnIleGlyLeuIleLeuSerGlyLysAla 139
 Db 527 AGTTATGTCGAACAAATCAAGAAACAAATTAATGACACATCTATGCTGTACCGCT 586
 QY 140 GlnLeuAsnGlnGlnIleLeuArgThrGlnGlyValProAlaGlnValHisAlaLeuMet 159
 Db 587 GAGATTAAACAAACAGTT-----GTAAATCTCTCCCTT 619
 QY 160 AspAsnGlyHisPheAlaAsnAspPrometArgTrpAsnGlnGlyTrpProTrpAsnAsn 179
 Db 620 GAT-----TCAAAAGCATTCATTAATACCAAGTAACCTTAACCTTAACCTTAAC 664
 QY 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
 Db 665 TTGACACCTGTTATTAATAAAGTAACCAAGGTGAACATCTTTTGTAGTCAACATGCA 724
 QY 191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTrpHisSerTrpProLeu 210
 Db 725 GCTACAGAGATGGTGTCTACTGCACTGCAATATATGAAATATCATATATACCTTAAC 784
 QY 211 GlnGlyGlnGlySerPheAspTrp-----HisAla 220
 Db 785 AAGGGTTGAAAGACTACACTTACACACTTAAGCTCAATTAACCATATTTCAACATCTCT 844
 QY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTrpAspTrpIleAsn--- 239
 Db 845 AAGAACTTG-----TTTGACGATATCTCTACTAGACATATCAACTGGAACAAACATC 895
 QY 240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTrpAlaThr 259
 Db 896 TTACTACTTATAGCGAAGAGAAATCTATCAAAAATG-----GCGATTTCAGAA 949

OY 300 LeuHISValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyLeu 319
 DB 1070 CAATCAACCGTAGGACTTACGAAACAGATGCGAAGCAATTAAGCAAAATTA 1129
 OY 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
 DB 1130 TCTCAAAACCAACAGATATACCAAGTGTGCGTAAAGTAGGCGAGATCCCTTGT 1189
 OY 340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyValSer 359
 DB 1190 ATCGATGCTGCTGACGAGCAGTACCTTACCACTGTTACGTGGGTGGGTGAGTCT 1249
 OY 360 AsnGlyPheTyrIleuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyValSer 379
 DB 1250 GACGCTCTTCCTCCGCTTACGAGCACTAACCTTACGCTTGTGACTGCTGGCGGCA 1309
 OY 380 IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394
 DB 1310 GCGGCTTCAACGCTTACCAAGTGTGTTAGCATCAACCT 1354
 AT 11
 LOCUS STSPBEC 1357 bp DNA linear BCT 18-MAY-2001
 DEFINITION Streptococcus pyogenes pyrogenic exotoxin B (speB) gene, complete cds.
 ACCESSION L26127.1 GI:431641
 VERSION L26127
 KEYWORDS
 SOURCE Streptococcus pyogenes.
 ORGANISM Streptococcus pyogenes.
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
 REFERENCE 1 (bases 1 to 1357)
 AUTHORS Kapur V., Topouzis S., Majesky M.W., Li L.L., Hamrick M.R., Hamill R.J., Patel J.M. and Musser J.M.
 TITLE A conserved Streptococcus pyogenes extracellular cysteine protease cleaves human fibronectin and degrades vitronectin
 JOURNAL Microb. Pathog. 15 (5), 327-346 (1993)
 MEDLINE 94285789
 PUBMED 7516997
 FEATURES
 source
 1. Location/Qualifiers
 1..1357
 /organism="Streptococcus pyogenes"
 /isolate="MGAS 289"
 /db_xref="taxon:1314"
 161..1357
 /gene="speB"
 161..1357
 /gene="speB"
 /function="cysteine protease precursor"
 /note="allele 12"
 /codon_start=1
 /transl_table=1
 /product="pyrogenic exotoxin B"
 /protein_id="AAA26993.1"
 /db_xref="GI:431642"
 Translation: "MNKKKIGRLISLALGGFVIANVPADONFARREKAKDSALT
 FIKSAAKAKARSAEDIKDKVNIIGGLSSNNYVITISGFIYSGDKRSPILG
 YSTGSDANKRENINASFMEIVEDIKKNKLDITTYATIKQPVVSLDSKIH
 NQGNPYNLITPVIEIKVPEQGSFVGQHAATCAVATQAIKYNHYPKRGKDYITL
 SSNNPYNHPNLPAISLTROYNNNNILPTVSGRESNOKAISETLADVGISVMDY
 GSSGSASSRVSORALKENFQSNVSHOINSDFSKDMESQIDKELSONPVYVQGV
 GKVGHAFFVIDAGDGRNRYHNMGKGVSDSDFRDLALNPALGTCGAGGPNQSA
 VVGIKP"
 BASE COUNT 447 a 263 c 266 g 381 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5,21e-24 Length: 1357
 Score: 496.00 Matches: 125
 Percent Similarity: 50.13% Conservative: 73
 Best Local Similarity: 31.65% Mismatches: 139
 Query Match: 11.18% Indels: 59

DB: 1 Gaps: 12
 US-10-030-330-1 (1-843) x STSPBEC (1-1357)
 OY 26 LysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnValSerLeuArgMetGly 45
 DB 267 AAGAAGCAAAAGATAGCCCTATCATATATCCAAAATACACA-GCATCAAGACGCT 325
 OY 46 GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnIleAspAlaGlu 65
 DB 326 GCACCAAGCGCAGAAAGAT--ATTAGCTTGCACAAAGATTACCTTAGCTGAGAACTT-- 379
 OY 66 ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
 DB 380 -----TCTGCTCTAATATGATATGATCTTACAAATTTCTTACTGCA 418
 OY 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
 DB 419 -----GGATTTGTTATCGTTTCAGAGATAAACGTTCCACAAATTTCTAGATAC 469
 OY 106 SerProIleGlyArgPheAspMetLysSerMetProAspAsnLeuArgMetIleLeuGln 125
 DB 470 TCTACCAAGCGATCATTTGACGCTTACCGT--AAAGAAAGATTGCTTCCATGAGA 526
 OY 126 IleTyr-----AspGlnGluIleGlyLeuIleLeuSerGlyLysAla 139
 DB 527 ACTTATGTGCAACCAATCAAGAAATTAATAATTAAGACACTTATCTGCTGACCGCT 586
 OY 140 GlnLeuAsnGluGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMet 159
 DB 587 GAGATTAAACACACAGTT-----GTTAATCTGCTCCTT 619
 OY 160 AspAsnGlyHisPheAlaAsnAspPheMetLysTrpAsnGlnGlyTyrProThrAsnAsn 179
 DB 620 GAT-----TCAAAAGCATTCATACCAACCAAGTAACCTTCAACCTTA 664
 OY 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
 DB 665 TTGACACCGCTTATTAAGAAATAAACCAGGTGAACATCTTTCTAGCTCAACATGCA 724
 OY 191 TyrThrGlyCysValAlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTrpLeu 210
 DB 725 GCTACAGCATGTGTTGCTCTACTCAACTGCTCAATATATGAATATCATATTAACCTAAC 784
 OY 211 GlnGlyGluGlySerPheAspTyr-----HisAla 220
 DB 785 AAGGAGTGAAGACTACCTTACACACTAAGCTCAATATTAACCATATTTCAACATCTCT 844
 OY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsn--- 239
 DB 845 AAGAACTTG-----TTTCAGCTATCTCTACTACAGCATATCAACCTGGAACACATC 895
 OY 240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAlaSerAlaTyrAlaThr 259
 DB 896 CTACCTACTTATAGCGGAAGAGAAATCTAAGCTTCAAAAATG-----CGCATTTTCAGAA 949
 OY 260 LeuMetLysAspValSerAlaSerValSerMetSerPheTyrGlnAsnGlySerGlyThr 279
 DB 950 TTGATGCGTGATGTTGATTTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1009
 OY 280 TyrSerValTyrValValGlyValLeuArgAsnAsnPheArgTyrLysArgSerLeuGln 299
 DB 1010 GGTACTCTCGCTTCAAAAGAGCTTGAAGAAACTTTGGCTACACCAATCTGTCTAC 1069
 OY 300 LeuHisValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyLeu 319
 DB 1070 CAATCAACCGTAGGACTTACGAAACAGATGCGAATCAACAATTAAGCAAAATTA 1129
 OY 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
 DB 1130 TCTCAAAACCAACAGATATACCAAGTGTGCGTAAAGTAGGCGAGATCCCTTGT 1189
 OY 340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyValSer 359

[illegible]

Db 326 GCACGAGGCCGACGAMAT ---ATTAAAGCTTGACAAAGTAACTTACGTAGGAGAACCTT--- 37
 Oy ArgGlyIleThrSerGlnGluGluGlySerProAlaIlePheTyrValAlaAsnArgGly 85
 Db 380 -----TCGGCTCTAATATGTATGTTTACAAATATTTCTACTGGA 411
 Oy 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
 Db 419 -----GATTTGTTATGCTTTCAGAGATTAAGCTTCGCCGAATATCTAGAGAT 465
 Oy 106 SerProIleGlyIleArgPheAspMetAspSerMetProAspLeuLeuArgMetTyrLeuGln 125
 Db 470 TCTACACCGGATCATTTGACGCTAACGCT ---AAGAAACATATGCTTCCTCATGGA 526
 Oy 126 IleTyr ---AspGlnGluIleGlyLeuIleLeuSerGlyTyrAla 139
 Db 527 AGTATATGTCGAACAAATCAAAAGAAACAAACAAATATTAACACTTATGCTGACGCAACGCT 586
 Oy 140 GlnLeuAsnGluGluIleLeuArgTyrThrGlnGlyValProAlaGluValAlaHisAlaLeuMet 159
 Db 587 GAGATTAACACCAACTT ---GTTAATCTTCCTT 619
 Oy 160 AspAsnGlyHisPheAlaAsnAspProweIleTyrPasnGlnGlyTyrProTyrPasnAsn 179
 Db 620 GAT-----TCAAAAGGCACTCATTTACACACCAAGCTAACCCCTTAACAACCTTA 664
 Oy 180 LysGluProLeuIleuProAsn-----GlyAsnHisAla 190
 Db 665 TTGCACACTCGTATATGAAAAAGTAAACCAAGCTGAACAACTTTTGTAGTGCATCATGCA 724
 Oy 191 TyrThrGlyCysValAlaIleThrAlaAlaAlaGlnIleMetArgTyrHisSerTyrProLeu 210
 Db 725 GCTACAGATGTGTGCTGCTCAAGCTCAACGCTCAAAATTAATCAATATCAATATTCACCTAAC 784
 Oy 211 GlnGlyGluGlySerPheAspTyr-----HisAla 220
 Db 785 AAGAGGTTGAAAGACTACACTTACACACTAACCTCAAAATCCAAATTTTCACCAACTCTT 844
 Oy 221 GlySerIleValGlyAsnTyrSerGlyThrPheGlyGluMetTyrAspTyrIleAsn--- 239
 Db 845 AAGAACTTG-----TTGCACACTATCTCTACATACAAATCAACTGGAACCAACATC 895
 Oy 240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
 Db 896 CTACACTACTATACCGGAAAGAAATCTAACCTTAACAAAAATG-----GCGATTTCAGAA 949
 Oy 260 LeuMetArgAspValSerAlaSerAlaSerMetSerPheTyrGluAsnGlySerGlyThr 279
 Db 950 TGTATGCTGATGTTGATTAATTCAGTACATGATTAATGCTCCATCTAAGTGGTTCTGCA 1009
 Oy 280 TyrSerValTyrValAlaGlyAlaLeuAlaArgAsnAsnPheArgTyrGlyArgSerLeuGln 299
 Db 1010 GGTACTCTCGTGTTCAAAAGGCTTGAAGAAACCTTGAGCAACCAACCAATCTGTTCCAC 1069
 Oy 300 LeuHisValArgAlaIleuTyrThrSerGlnGluTyrHisAspMetIleArgGlyGlnLeu 319
 Db 1070 CAATATCAACCGCTACGCACTTTACCAAAAGAAATGGAATGCAAAATTTGACAAAGAAATTA 1129
 Oy 320 AlaSerGluArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
 Db 1130 TCTCAAAACCAACCACTATATACCAAGGTGTCGTTAAAGTAAAGGACGACATGCTTGTGT 1189
 Oy 340 CysAspIleTyrAlaSerAspIleThrPheHisPheAsnTyrProGlyTyrGlyValSer 359
 Db 1190 ATCGATGTGGCTGACGACGACTTCTTCAACATGTTAACTGGGCTGGGGTGAAGCTCT 1249
 Oy 360 AsnGlyPheTyrLeuLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGluGlyLeu 1250
 Db 1250 GACGCTTTCTCCGTTGACGCACTAAACCTTCAGCTTGTGATCTGCTGGCGGCA 1309
 Oy 380 IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394

DB 1310 GCGGCTTCAACGGTTACCAAGTCTGTGTGAGCATCAACCT 1354

RESULT 13
STRSPEBK 1357 bp DNA linear BCT 18-MAY-2001
LOCUS Streptococcus pyogenes pyrogenic exotoxin B (speb) gene, complete
DEFINITION cds.
ACCESSION L26135
KEYWORDS L26135.1 GI:431657
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 1357)
Kapirya, V., Topouzis, S., Majesky, M.W., Li, L.L., Hamrick, M.R.,
Hamill, R.J., Patel, J.M. and Musser, J.M.
A conserved Streptococcus pyogenes extracellular cysteine protease
cleaves human fibronectin and degrades vitronectin
Microb. Pathog. 15 (5), 327-346 (1993)
JOURNAL
LINE
94285789
SUBMED
7516997

FEATURES
source
Location/Qualifiers
1..1357
/organism="Streptococcus pyogenes"
/isolate="MGAS 796"
/db_xref="taxon:1314"
161..1357
/gene="speb"
161..1357
/gene="speb"
/function="cysteine protease precursor"
/note="allele 19"
/codon_start=1
/transl_table=11
/product="pyrogenic exotoxin B"
/protein_id="AA27001.1"
/db_xref="GI:431658"
/translation="MNKKIGIRLISLILAGFVLNPFVADNPNRNEKESKDSAIT
FIKSAATKAKARBAEDIKDKVNIAGGLSSNNMYVISTGFIYSGDKSPETLG
YSTGSPFNKRENTASFWESVEQIKENKLDVTYAGTAETKQVNVSLDSKGIHY
NQGPNFLTPVIEKVPKQSEFVQGHATCVAATQIMKYNHPKRGKLDYVTL
SSNPYFNHPNLTALISTROYNNNNIIPITSGRESNPKQKASISLADVGISVMDY
GSSGSASSRSVQALKENGFNOVHOINSDFSKQMEAOIDKELSONOVYVGV
GKVGHAFFVIDGADGRNFYHVMWGMGVSDGFPRDALNPALSTGTGGAGGNGVQSA
VVGTKP"

BASE COUNT 448 a 261 c 266 g 382 t

Alignment Scores:
No.: 5.21e-24 Length: 1357
Score: 496.00 Matches: 125
Percent Similarity: 50.388 Conservaive: 74
Best Local Similarity: 31.658 Mismatches: 138
Query Match: 11.18% Gaps: 59
DB: 1 12

US-10-030-330-1 (1-843) x STRSPEBK (1-1357)

QY 26 LysgluAArgAlaLeuSerLeuAlaArgLeuAlaLeuArgAlaValSerLeuArgMetGly 45
DB 267 AAGAGCAAAAGATAGCGCTATACATTATCCAAATAACAGCA-GCATCAAAACAGGT 325
QY 46 GluThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu 65
DB 326 GCACGACGCGCAGAGAT--ATTAGCTTGACAAAGTTAACTTAGTGAGCAACTT-- 379
QY 66 ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
DB 380 -----TCGCGCTCTAATATGATGTTTACAATATTCTTCTACTGGA 418
QY 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgGlyProThrIleLeuAlaTyr 105
||||| :|||: ||| ||| ||| ||| |||

DB 419 -----GAAATTGTATCGTTTCAGAGATTAACGTTCTCCAGAAATTCAGATAC 469
QY 106 SerProIleGlyArgPheAspMetLaspSerMetProAspAsnLeuArgMetIleGlu 125
DB 470 TCTACAGCGGATCATTTGATGTAAACGTT--AAAGAAACATCTCTTCATCAAGAA 526
QY 126 IleTyr-----AspGlnIleGlyLeuIleLeuSerGlyIleVal 139
DB 527 ACTTATGTCGAACAAATCAAGAAACAAATAATTAGACACTTATATCTGCTGACCGGT 586
QY 140 GlnLeuAsnGluGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMet 159
DB 587 GAGATTAAACACCAAGTT-----GTAAATCTCTCCCT 619
QY 160 AspAsnGlnHisPheAlaAsnAspProMetLgTyrPaspGlnGlyTyrProTyrAsnAsn 179
DB 620 GAT-----TCAAAAGCATTCATTACACCAAGTAACTTCAACCA 664
QY 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
DB 665 TTGACACCTGTTATTTGAAAAAGTAAACCAAGGTGAACAAATCTTTGTAGTCAACTGCA 724
QY 191 TyrThrGlyCysValAlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTyrProLeu 210
DB 725 GCTACAGGATGTGTGCTACTGCAACTGCTCAATATATGAAATATCATATATCACTTAAT 784
QY 211 GlnGlyGluGlySerPheAspTyr-----HisAla 220
DB 785 AAAGGTTGAAAGACTCACTTACACACTTAAGCTCAATTAACCAATATTTCAACACTCT 844
QY 221 GlySerLeuValGlyAsnThrSerGlyThrPheGlyGluMetTyrAspThrIleAsn---- 239
DB 845 AAGAACTTC-----TTTGACAGTACTCTCTACTAGCAATCACTGGAACACATCCT 895
QY 240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
DB 896 CTACTACTTATACGGAGAGAGATCACTTCAAAAATG-----CGAATTTCAGAA 949
QY 260 LeuMetLgAspValSerAlaSerValSerMetSerPheTyrGluAsnIleSerGlyThr 279
DB 950 TTGATGCGCTGATGTTGTTTTCAGTACAGCATGATGATGATGATGATGATGATGATGAT 1009
QY 280 TyrSerValTyrValValGlyAlaLeuAlaGlnAspAsnPheArgTyrLysArgSerLeuGln 299
DB 1010 GGTACTCTCGGTTCAAGACCTTGAAGAAATCTTTGGCTACCAACCAATCTGTTCAC 1069
QY 300 LeuHisValArgAlaLeuTyrThrSerGlnGluTyrPheHisAspMetIleArgGlyIleLeu 319
DB 1070 CAATCAACCGTAGTACTTTAGCAAAACAGATTGGGAAGACAAATTTGCAAAAGAAATTA 1129
QY 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
DB 1130 TCTCAAAACCAACGATATATCAAGAGGTGCGGTAAAGTACGCGACATGCTTGTGTT 1189
QY 340 CysAspGlyTyrAlaSerAspLysIleThrPheHisPheAsnThrPglTyrPglGlyValSer 359
DB 1190 ATCGATGTCGTCGACGAGCACTAATCTTACATGTTAACTGAGGTTGGGAGTCTCT 1249
QY 360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGly 379
DB 1250 GATGCTCTCTTCGCTTACCCACCAACCCCTTTCAGCTGTGCTGCTGCTGCTGCTGCTG 1309
QY 380 IleGlyPheThrIleTyrGlnIleIleThrGlnGlyIleGluPro 394
DB 1310 GCGGCTTCAACGGTTACCAAGTCTGTGTGAGCATCAACCT 1354

RESULT 14
STRSPEBK 1357 bp DNA linear BCT 18-MAY-2001
LOCUS Streptococcus pyogenes pyrogenic exotoxin B (speb) gene, complete
DEFINITION cds.
ACCESSION L26136
VERSION L26136.1 GI:431659

CDS

161..1357
 /gene="speb"
 /function="cysteine protease precursor"
 /note="allele 30"
 /codon_start=1
 /transl_table=11
 /product="pyrogenic exotoxin B"
 /protein_id="AA27013.1"
 /db_xref="GI:431682"
 /translation="MNKKKLGIRLLSLALGGFVLANPYVADONFARNREKAKSDAIF
 /OKSAATIKAGKRSHEIDIKLDVNLGGELSGSNMIVYINISGFEVYSGDKSPETLG
 YSTSGSFVNGKENIASFMESVEEQIKENKILDTYTAETAEIKQPVVSLDSSKIH
 NOGNYNLTPVIEKVPGEQSFVQHAATGCVATAQIMKRYHNPYKMGIDYTL
 SSNNPYFNHPKILFALISTROYNNNNIIPYSGRESNTOKMAISLMDVIGISMDY
 GPSSGASSRYVQALKEKNGFYNGSHOINRSDFSKODMEQIDKELSONPYYQGV
 GKVGSHAFVVDGADGRNRYHVMWNGCVSDGFRDIALNPSALGTGGCAGGNGVQSA
 VVGIRP"

BASE COUNT 448 a 264 c 266 g 379 t

Comment Scores:

Score: 5.21e-24 Length: 1357
 Percent Similarity: 496.00 Matches: 125
 Best Local Similarity: 50.38% Conservative: 74
 Query Match: 31.65% Mismatches: 138
 11.18% Indels: 59
 Gaps: 12

US-10-030-330-1 (1-843) x STRSPEBW (1-1357)

QY 26 LysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnValSerLeuArgMetGly 45
 DB 267 AAGAAGCAAAAGATGCGCTATCATTTATCCAAAATACAGCA-GCTATCAAGACAGGT 325
 QY 46 GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu 65
 DB 326 GCACGACGCGCAGAGAT--ATTAGCTTGACAAAGTTAACTTAGCTGACGACAACTT-- 379
 QY 66 ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
 DB 380 -----TCTGGCTCTATATATGATATGATTTTACAAATTTCTTACTGGA 418
 QY 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
 DB 419 -----GCATTGTTGTTTCCTTACAGAGATAACGTTCCAGAAATTTCTAGGATAC 469
 QY 106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGln 125
 DB 470 TCTACCAAGCGGATTTGACGTTAACGGT---AAGAAACAATTCCTTCTCATGAGA 526
 QY 126 IleTyr-----AspGlnGluIleGlyLeuIleLeuSerGlyLysAla 139
 DB 527 ACTTATGTCGACAAATCAAGAAACAAATAATGACACACTTATGCTGTACCGGT 586
 QY 140 GlnLeuAsnGluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMet 159
 DB 587 GAGATTAAACAACCAAGCTT-----GTTAAATCTCTCCTT 619
 QY 160 AspAsnGlnHisPheAlaAsnAspPheMetArgTrpAsnGlnGlyTyrProTrpAsn 179
 DB 620 GAT-----TCAAAAGGCAATTCATTACAAACCAAGTAAACCTTACAAACCTTA 664
 QY 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
 DB 665 TTGACACCTGTATTGAAATAAAGTAAACAGGTGAACAATCTTTGTAGTCAACATGGA 724
 QY 191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTrpProLeu 210
 DB 725 GCTACAGATGTGTGCTACTGCAACTGCTCAAAATTATGAAATATCATATTAACCTTAC 784
 QY 211 GlnGlyGluGlySerPheAspTyr-----HisAla 220
 DB 785 AAGGGTTGAAGAAAGACTTACACACTAAGCTCAATATTAACCAATATTTCACACATCTCT 844

QY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsn--- 239
 DB 845 AAGAAGCTTG-----TTTGCAGCTATCTCTACTAGACATATACACTGGAACAACATC 895
 QY 240 MetProGlyAsnProAspLeuAspLeuAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
 DB 896 TTACCTACTTATAGCGGAAGAGAACTTACGTTCAAAAAAATG-----GCCATTTCAGAA 949
 QY 260 LeuMetAlaGAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThr 279
 DB 950 TTGATGCGCTGATGGTGGTATTTCACTAGACATGATTAATGTCATCTAGTGGTCTGCA 1009
 QY 280 TyrSerAlaTyrValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGln 299
 DB 1010 GGTACTCTCGGTTCCAAAGACCTTGAAGAAACCTTGGCTACACCAATCTTCTTAC 1069
 QY 300 LeuHisValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyLeuLeu 319
 DB 1070 CAATCAACCGTAGGAGCTTATGCAACACMAATGGAAGCACAATTCACAAAGAAATTA 1129
 QY 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
 DB 1130 TCTCAAAACCAACAGATATACCAAGGTGTCGTTAAAGTACGCGGACATCCCTTTGT 1189
 QY 340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyLysArgSer 359
 DB 1190 ATCGATGCGTGCAGAGGACGATACCTTACATGTTACTGCGGTGGGTGGAGCTCT 1249
 QY 360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGly 379
 DB 1250 GACGGCTCTCTCCGTTTCGACGACCTAAACCTTCAGCTCTGTGACTGTCGCGCGCA 1309
 QY 380 IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394
 DB 1310 GCGGCTTCACGCTTACCAAGTCTGTTGATGACATCAAACT 1354

RESULT 17
 STRSPEBW 1357 bp DNA linear BCT 18-MAY-2001
 LOCUS 1357 bp DNA linear BCT 18-MAY-2001
 DEFINITION
 accession L26149.1 GI:431685
 version 1
 keywords
 source Streptococcus pyogenes.
 organism Streptococcus pyogenes.
 Streptococcus pyogenes.
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 1 (bases 1 to 1357)
 Kapur,V., Topouzis,S., Majesky,M.W., Li,L.L., Hamrick,M.R.,
 Hamill,R.J., Patel,J.M. and Musser,J.M.
 A conserved Streptococcus pyogenes extracellular cysteine protease
 cleaves human fibronectin and degrades vitronectin
 Microb. Pathog. 15 (5), 327-346 (1993)
 MEDLINE 94285789
 PUBMED 7516997

FEATURES
 source
 gene
 CDS
 1..1357
 /organism="Streptococcus pyogenes"
 /isolate="MGAS 1898"
 /db_xref="taxon:1314"
 161..1357
 /gene="speb"
 /gene="speb"
 161..1357
 /function="cysteine protease precursor"
 /note="allele 32"
 /codon_start=1
 /transl_table=11
 /product="pyrogenic exotoxin B"
 /protein_id="AA27015.1"
 /db_xref="GI:431686"

/bunsallation="MNKKRIGJLILSLAGVLPANPFAVDPAFNRNKRKAPASATIT
FIOKSAALIKGARGAEADIKDNRNVLGCEGSGSMRYVNTSTGFEFVAGDSREIIL
YGTSGSDPANGKEIKEIFAFEMYSVEOATKEKKKATDITAGYAGAEIKOVVPSILDSNG
NOSTGSDYNLFLVEIKERFVGGOSFVGOHAKKATCAVATTAOIMKYHNPYKKGADTYIT
SNMNPYNHNPENLFAALSTGYMNWNLITPSYRESBNOKAMISELSDNDVSDYMDY
GPGSGAGSGSVORALAKENGYNOVHOALINRSPFSDWDESOIIMELSONDPVYOGG
KRGVGHAFVFDGADGRNFYHVNMGMGSGVSDGDFRDLALNPMLAGCGGAGGFCNGSOA
VVGIC

BASE COUNT	448 a	263 c	265 g	381 t
ORIGIN			.	

Alignment Scores:	
Pred. No.:	5,21e-24
Score:	496.00
Percent Similarity:	50.13%
Best Local Similarity:	31.65%
Query Match:	11.18%
OB:	1
Length:	1357
Matches:	125
Conservative:	73
Mismatches:	139
Indels:	59
Gaps:	12

[illegible]

Db	1190	ATCATGCTGCTGACGACGACGTAACCTTCTACCATGTATTACCTGGGGTGGGGAGATTCT	1249
Oy	360	AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGly	379
Db	1250	GACGGCTCTCCGCTCTGACGACGACGTAACCTTCAGCTCTTGCTAGCTAGCGTGGCGGCCA	1309
Oy	380	IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro	394
Db	1310	GGCGGCTTCAACGGGTTACCAAAAGTCGTGTAGACGATCAAAAGCT	1354
RESULT 20			
STRSPEBH			
LOCUS	STRSPEBH	1357 bp	DNA linear BCT 18-MAY-2001
DEFINITION	Streptococcus pyogenes pyrogenic exotoxin B (speb) gene, complete cds.		
ACCESSION	L261332		
VERSION	L261332.1	GI:431651	
KEYWORDS			
SOURCE	Streptococcus pyogenes.		
ORGANISM	Streptococcus pyogenes		
REFERENCE	Bacteria: Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
AUTHORS	1 (bases 1 to 1357) Kapuri,V., Topouzis,S., Majesky,M.W., Li,L.L., Hamrick,M.R., Hamill,R.J., Patti,J.M. and Nussler,J.M.		
TITLE	A conserved Streptococcus pyogenes extracellular cysteine protease cleaves human fibronectin and degrades vitronectin		
JOURNAL	Microb. pathog. 15 (5), 327-346 (1993)		
MEDLINE	94285789		
PUBMED	7516997		
FEATURES			
source	Location/Qualifiers 1..1357 /organism="Streptococcus pyogenes" /isolate="MGAS 427" /db_xref="taxon:1314" 161..1357 /gene="speb" 161..1357 /gene="speb" /function="cysteine protease precursor" /note="allele 16" /codon_start=1 /transl_table=1 /product="pyrogenic exotoxin B" /protein_id="AA26998.1" /db_xref="GI:431652"		
gene	/translation="MKKKKIGRLSLAAGFYLANPVPADQFARNEKRSKSAITTFIKSAIKAGARSAEDIKIDKVLVAGELSGSNMYVINISTGCVIVSGDRSPETLIGYSTGSDPANGKENIASFMEYSYVDIENKRLDITVAGTAEIKQPVKSLDSKGIHYNOGPNYLTLPVIEKVRGEOSFYGOAAATGCVATAQIKRYNPNKGVDYTYTLSSRNPNHPRNLEPAALSTROYNNNTILPYSGESVNOVKMAISELMQNDISMDYGVGRSGSGSSRYQALSTENGYCNSQVNIHRSPDSKDDWEQKIDKELSQNDPYQYQVGRVGHAFVVIDGADGRNRFHYHNMGMGVSDGFRLDLNPSALGTGGAGGFGNYQSAVVGAKP"		
CDS			
BASE COUNT	448 a	265 c	267 g 377 t
ORIGIN			
Alignment Scores:			
Pred. No.:	6.06e-24	Length:	1357
Score:	495.00	Matches:	125
Score Similarity:	50.13%	Conservative:	73
Best Local Similarity:	31.65%	Mismatches:	139
Query Match:	11.15%	Indels:	59
Db:	1	Gaps:	12
US-10-030-330-1 (1-843)	x	STRSPEBH (1-1357)	
Oy	26	LYSGIURGALASERLSEUAIARGLEUALSEUARGINVALSERLEUARGMETGLY	45
Db	267	AAGAAGCAAAAGATAGCGCTATTCACATTATCCAAAATCGCA-GCATCAAAAGCAGT	325
Oy	46	GIATHRALVALSERASPLYSILESERILEASPTYRVALTYRARGINGLYASPALAGLU	65

```

Db      326 GCACGAGCGGAGAGAT---ATTACCTTGACAAATTAAGTGGAGACTT--- 379
Qy      66 ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
Db      380 -----TCGCGCTTAATGTATGTATGTATACATATTCTTACTGGA 418
Qy      86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
Db      419 -----GGATTGTGTTATCGTTTCAGAGATTAACGTTCTCCAGAAATTCAGATAC 469
Qy      106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGln 125
Db      470 TCTACACGCGGATCATTTGACGCTAACGCT---AAGAAATAACATTGCTTCTTCATGGAA 526
Qy      126 IleTyr-----AspGlnGluIleGlyLeuIleLeuSerGlyLysAla 139
Db      527 AGTTATGTGACAAACAAATCAAGAAACAAATAATAGACACTATTATGCTGCTACCGCT 586
Qy      140 GlnLeuAsnGluGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMet 159
Db      587 GACATTAAACACACAGTT-----GTTAATCTCTCCTT 619
Qy      160 AspAsnGlyHisPheAlaAsnAspPrometArgTrpAsnGlnGlyTyrProTrpAsnAsn 179
Db      620 GAT-----TCAAAAGCATTCATTACATCAAGTAAACCTTACAAACCTA 664
Qy      180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
Db      665 TTCACACCTGTTTATTAATAAAGTAATAACAGGTGAACAATCTTTGTAGCTCAACATGCA 724
Qy      191 TyrThrGlyCysValAlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTrpProLeu 210
Db      725 GCTACAGGATGTGTTCTCTACTGCACTGCTCAATATGAAATATCATATACCTTAAC 784
Qy      211 GlnGluGluGlySerPheAspTyr-----HisAla 220
Db      785 AAAGGGGTGAAGACTACACTACACTACACTACATTAATTAATTTCAACATCTCT 844
Qy      221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsn--- 239
Db      845 AAGAACTTG-----TTTCAGCTATCTCTACTAGACATACAACTGGAAACAACATC 895
Qy      240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
Db      896 CTACCTACTTATAGCGGAGAGATCTTAACGTTCAAAAAATG-----GCGATTTCAGAA 949
Qy      260 LeuMetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThr 279
Db      950 TTGATGGCTGATGTTGTTGATTTACATAGACATGATTAAGTTCATCTAGTGTCTGCA 1009
Qy      280 TyrSerValTyrValAlaGlyAlaLeuArgAsnAsnPheArgTyrGlyArgSerLeuGln 299
Db      1010 GGTAGCTCTCGTCTCAAGAGCCTTGAAGAAACTTTGGCTACAAACCATCTGTTCAAC 1069
Qy      300 LeuHisValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyLeuLeu 319
Db      1070 CAATCAACCGTAGGACTTTAGCAACAAAGATTGGGAGACCAAAATTCACAAAGAAATTA 1129
Qy      320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
Db      1130 TCTCAAAACCAACCAAGTATCTACCAAGTCTCGGTAAAGGCGGACATGCTTGT 1189
Qy      340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTrpGlyValSer 359
Db      1190 ATCATGTGTGCTGACGACGACGATTAAGTCAATGTAAGTGGGTGGGAGTCTCT 1249
Qy      360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGly 379
Db      1250 GACGCGCTTCTTCGCTTTCAGACCACTTAACCTTACCTTGTGTTGCGGCGCA 1309
Qy      380 IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394

```

```

Db      1310 GCGGCTTCACGCTTACCAAAAGTGTGTTGAGGCATCAAACT 1354
Search completed: June 2, 2003, 00:13:31
Job time : 5689 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 1, 2003, 22:27:31 ; Search time 469 seconds

(without alignments)
4047.835 Million cell updates/sec

Title: US-10-030-330-1

Perfect score: 4438
Sequence: 1 MKSFLAIVLFGIAGQH.....PNCATILKVDGTTTINIVH 843

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Command line parameters:
-MODE=frame_p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US1003030/runat.23052003.181643.6744/app.query.fasta.1.1031
-DB=N.Geneseq.101002 -OFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -FRAN=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=20
-MODE=LOCAL -OUTFMT=ptio -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US1003030.ecgn.1.1.290.etrnat.23052003.181643.6744 -NCPU=6 -ICPU=3
-NO_XIPX -NO_MMAP -LANG=EUENT -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -NARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq.101002: **
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4438	100.0	2529	20	AAK91703	Porphyromonas ging
2	4438	100.0	2532	21	AAK67909	Porphyromonas ging
3	4438	100.0	2537	20	AAK91578	Porphyromonas ging
4	493	11.1	1194	24	ABN70210	Streptococcus poly
5	484	10.9	1197	17	AAK45219	Streptococcus pyog
6	484	10.9	1197	17	AAK15294	S. pyogenes spee e
7	484	10.9	1197	21	AAK07111	S. pyogenes cyste
8	331.5	7.5	415	21	AAK82217	N. meningitidis pa
9	190	4.3	1620	20	AAK91791	Porphyromonas ging
10	190	4.3	1689	20	AAK91648	Porphyromonas ging
11	164	3.7	2463	20	AAK91597	Porphyromonas ging
12	164	3.7	2607	20	AAK91572	Porphyromonas ging
13	159.5	3.6	2763	20	AAK91740	Porphyromonas ging
14	159.5	3.6	2766	20	AAK91739	Porphyromonas ging
15	159.5	3.6	2775	20	AAK91738	Porphyromonas ging
16	159.5	3.6	2814	20	AAK91610	Porphyromonas ging
17	154	3.5	3978	22	AAK81312	Escherichia coli p
18	148.5	3.3	4059	22	AAK75059	Nucleotide sequenc
19	148.5	3.3	9047	20	AAK12949	Enterococcus faeca
20	148.5	3.3	9466	23	AAK75834	DNA encoding novel
21	147.5	3.3	3413	20	AAK72330	F. dalsuvarum CP70
22	144.5	3.3	4986	24	AAK37774	Alpha-tomallotryl-
23	144	3.2	1163020	24	ABO67197	Listeria innocua c
24	144	3.2	2944528	24	ABA03041	Listeria monocyog
25	144	3.2	3011208	24	ABO69245	Listeria innocua d
26	143	3.2	1170	20	AAK91662	Porphyromonas ging
27	143	3.2	1173	20	AAK91541	Porphyromonas ging
28	143	3.2	1725	24	ABO69158	Listeria monocyog
29	143	3.2	2421	24	ABO70811	Listeria monocyog
30	142.5	3.2	1830121	17	AAK42063	Haemophilus influe
31	142.5	3.2	2823	24	ABO69232	Listeria sp DNA se
32	140	3.2	3106	18	AAK99805	Enterobacteriaceae
33	140	3.2	6159	22	AAK59104	M. catarrhalis str
34	140	3.2	30349	24	ABN90859	Staphylococcus epi
35	139.5	3.1	2848	18	AAK86089	Kex2 protease enco
36	139.5	3.1	3091	10	AAK90387	Gene encoding KEX2
37	139.5	3.1	4106	13	AAK98696	Yeast KEX2 gene.
38	139.5	3.1	4118	16	AAK029472	Muramidase release
39	139.5	3.1	9475	23	AAK568878	DNA encoding novel
40	139	3.1	3509	24	ABN92640	Blifidobacterium bi
41	139	3.1	9432	17	AAK09312	Lkp1 operon. Haem
42	138.5	3.1	2809	16	AAK51531	Mycobacterium gall
43	138.5	3.1	8709	22	AAK25628	H. pylori HPN165 e
44	138	3.1	3106	18	AAK02323	Enterohaemorrhagic
45	138	3.1	4765	17	AAK08132	Protease gene. Py

ALIGNMENTS

RESULT 1
AAK91703
ID AAK91703 standard; DNA: 2529 BP.
XX
XX
XX
AC AAK91703;
XX
XX
DT 25-AUG-1999 (first entry)
XX
XX
DE Porphyromonas gingivalis protein PG28 encoding DNA.
XX
XX
KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
XX
XX
XX vaccine; antigenic; ds.
XX
XX
OS Porphyromonas gingivalis.
XX
XX
PN W09929870-A1.
XX
PD 17-JUN-1999.

XX 10-DEC-1998: 98W0-AU01023.
 XX 04-AUG-1998: 98AU-0005028.
 PR 10-DEC-1997: 97AU-0000839.
 PR 31-DEC-1997: 97AU-0001182.
 PR 30-JAN-1998: 98AU-0001546.
 PR 10-MAR-1998: 98AU-0002264.
 PR 09-APR-1998: 98AU-0002911.
 PR 23-APR-1998: 98AU-0003128.
 PR 05-MAY-1998: 98AU-0003338.
 PR 22-MAY-1998: 98AU-0003654.
 PR 29-JUL-1998: 98AU-0004917.
 XX (CSLC-) CSL LTD.
 XX Agnes CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rother LJ, Webb EA;

WPI: 1999-385613/32.
 P-PSDB: AAY34485.

PT Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 PS Claim 12: Page 209-210; 588bp; English.

CC AAY91536 to AAY91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY91538 to
 CC AAY91581. AAY91802 to AAY91899 represent PCR primers used in the
 CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The Pg polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.

SQ Sequence 2529 BP; 722 A; 561 C; 608 G; 638 T; 0 other;

Alignment Scores:
 Pred. No.: 0 Length: 2529
 Score: 4438.00 Matches: 843
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 20 Gaps: 0

10-030-330-1 (1-843) x AAY91703 (1-2529)

1 MetLysLysSerPheLeuLeuAlaIleValMetLeuPheGlyIleAlaMetGlnGlyHis 20
 1 ATGAAATAAAAGTTTCTTTTACCAATGATGCTCTTGGCATTCATGACAGGACAT 60
 21 SerAlaProValThrIleGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgIleVal 40
 61 TGTGCTCCGCTTACGAAAGAGAGCTTTGAGCTGCTGCGCTTGGACAGGATA 120
 41 SerLeuArgMetGlyGlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArg 60
 121 TCTTTCGCAATGGGCAACACAGCAGTATGTGACACAGATTTCATGATTCATGCTTATCGG 180
 61 GlnGlyAspAlaGluArgGlyIleThrSerGlnGluGluGlySerProAlaTyrPheTyr 80
 181 CAAGGAGATGCTGAGAGGGGTATCATCATCAAGAGGAGGCTCTCTCATATTTTAT 240
 81 ValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuValAlaAlaAspArgIlePro 100
 241 GTAGCTAATCTGTAATATGAGAGGCTATGCTTGTAGCAGCAATATGACAGAAATCCG 300
 101 ThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAspAlaLeu 120
 301 ACAATTTTAGCTAATCCACCATGCTGCTTTCGACATGACAGTATGCGGACCAATCTT 360

QY 121 ArgMetTrpLeuGlnIleTyrAspGlnGluIleGlyLeuIleLeuSerGlyLysAlaGln 140
 DB 361 CCGATCTGGCTACAAATTTACGATCGAAGAAATAGGCTATCTTCGCGAAAGCTCAG 420
 QY 141 LeuAsnGluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMetAsp 160
 DB 421 CTCATGAGAGATATATACGATCCGAGGCGTACCCGCTGAAAGTATCATCTCTGATGAT 480
 QY 161 AsnGlyHisPheAlaAsnAspProMetArgTyrAspGlnGlyTyrProIleAsnLeu 180
 DB 481 AACGGTCATTTTGGCAACGATCCCATGGAATCAAGGTTACCATGGAACCAATTAAG 540
 QY 181 GluProLeuLeuProAsnGlnAsnHisAlaTyrThrGlyCysValAlaThrAlaAla 200
 DB 541 GAACCATGCTCTCTAATGGAATCATATGCTATACCGGCTGTCTCTCTGTCGAGA 600
 QY 201 GlnIleMetArgTyrHisSerTrpProLeuGlnGlyGlySerPheAspTyrHisAla 220
 DB 601 CAATCATCTGCTACCATAGCTGCGCTTCAAGGTGAAGGCTCTTTCATATATATGCA 660
 QY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsnMet 240
 DB 661 GGTTCATTAGTTGGCAACTGTCCGCAATTTGGTGAATGTACAGTGCATCATATATG 720
 QY 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAlaAspAlaTyrAlaThrLeu 260
 DB 721 CCGGAAATCCGACCTTATATCATGACTCAATCATGATGATGATGATGATGATGATGATG 780
 QY 261 MetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
 DB 781 ATGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
 QY 281 SerValTyrValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeu 300
 DB 841 AGCGTTATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 QY 301 HisValArgAlaLeuTyrThrSerGlnGluTyrPheAspMetIleArgIleGluLeuAla 320
 DB 901 CATGTACGCGCTTATATATCATGACTCAATCATGATGATGATGATGATGATGATGATGATG 960
 QY 321 SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCys 340
 DB 961 TCCGGAAGGCGGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
 QY 341 AspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrPheGlyValSerAsn 360
 DB 1021 GATGCTATGCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
 QY 361 GlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGlyIle 380
 DB 1081 GGTCTCTACAACTAATCATCTCTGCGGACCTTCTGCGGATCGAGTGAAGGAGATA 1140
 QY 381 GlyPheThrIleTyrGlnGluIleIleThrGlyIleGluProAlaLysThrProAlaGlu 400
 DB 1141 GGTTTTACCATTTATCAAGATCATCACCGGATGAAACCGCTAAGCTCCCGGTGAA 1200
 QY 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluAlaGlyTyrLysSer 420
 DB 1201 GCGGATGAGATGCTTGGCGATCTTGGCAGTACGAAAGATGAAAGCGAGATATAAAT 1260
 QY 421 GluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGluGluGlnSerAsnLeu 440
 DB 1261 GAATCGGATGTAAGCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
 QY 441 AspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluValIleGluValLysThrSerSer 460
 DB 1321 GACCTCGATACAGATGTAACAAAGGCTGACGGAAGTATGATGATGATGATGATGATGATG 1380
 QY 461 IleAsnIleSerTrpTyrGlyTyrGlyGluHisProGluLysPheSerLeuAlaProAsn 480
 DB 1381 ATCAATATCTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
 QY 481 GlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgThrGlyThrGluGln 500

```

DB      1441 CAGTTGTCACAGAGATCAACACCACTACTTATTCGCGCACAGCCGACAG
QY      501  TTPGIUPROVALARHISALAGINLYGYTYRVALASERILEYVALASNTHT 520
DB      1501 TGGAGCCGGTAGCGGACGACAGAGAGATATGTCAAATGAATTAATAATACGACA
QY      521  ASPPROASNAVALVALVALTHRVALASPAASNGILYLSLEUSERILEVALPRO 540
DB      1561 GACCCGAAACATGTGTACTACGCTAGTAAATTAACGAAGGCAAGCTCATTCGTC
QY      541  ASNSERPHEVALALASPLEUASNSERTYRGUHSISERTHTLETHRVALGLNPHASN 560
DB      1621 AACAGCTTTGTCGAGATCTGAATTCATGAACATAGACATTAACAGTTCAT
QY      561  SERASPSERPROASGUILLEARGTHRPROVALALAPHEALUSERHTGLYALATHR 580
DB      1661 AGCGACACCCCTGAGAGAGATCGTACACCCGTCGCTTCTATCTACAGAGCTACT
QY      581  ALAASPAPVALILESERLEUGLYTRPVALMETALAGLVALPROGLYGLYSERASN 600
DB      1741 GCGGACGATGTAATCTTTGGGCTGGGTAAGCTGGAAGTTCCGGGCGGTACGAC
QY      601  TYRPROVALVALTRPSERLYSASPVALLEUTHRLEUSERGLYLASPTYRTHLEUPR 620
DB      1801 TATCCGTTGGTTGGTCTAAAGACGTTCTCACTCTCGGAAGGCGACTATACATGTCG
QY      621  TYRARGPHERILEASASNGILNYSAPGILUTPHYLSYILEGLYSERVALSERVAL 640
DB      1861 TATAGATTTTCCATCAACACCAAAAGATGAATGAATGAATGAATGAATGATGAT
QY      641  LYSTHRPROTHGLUTYRTHRHSISPROLEUPHEGLUVALGLYHSASNGILNTHSE 660
DB      1921 AAAACACCGACGAGATATACGACCCCTTATTCGAAGTGGCCATTAATACTTACC
QY      661  TYRTHRLEUASPMETALHISASNARGVALLEUPROASPHERTHRLEULYSANLEUG 680
DB      1981 TATACGTCGATATGCGACACACAGAGATATGCCGACTTACACTCAAAATCTCGSA
QY      681  LEUPROPHASNGILYLULEUVALVALPHEARGINLTHGLISERSESGLYSER 700
DB      2041 TTGCTTTCAANGGAGCTGTGGTGTGTTTCCGCAACCAATCCCATCGGGGCTT
QY      701  LEUTRPAALAGINLUTHRVALHISILEYSGINGLYGLUTHRPHRVALYRILYSPRO 720
DB      2101 TTATGCGAGCTCAGAAACAGTACATATCAAGCAAGGAGAACTTCTATATTAACCT
QY      721  VALVALGLUGLYPROILEPROASPGILYSEPTYRARGALATHRLEUHSIALAPHEVAL 740
DB      2161 GTTGTGGAAGGCCCTATACCTGATGCCATCGTATCGTGACCCCTCATGTCGTAAC
QY      741  GLYINGINGINLUTHRLEULYSGLYLSATGAANTYRTHVALYLSILEVALSNGLY 760
DB      2221 GGACAAACACAGTGTGTACTCTCAAGGGGAAAGAACTACGCTGTAAGTCTCATAGT
QY      761  THRALAVALGLUALALEGLUSERSERGLUGILYLEARGVALPHEPROASNPROMAARG 780
DB      2281 ACAGCGGTAGAACATAGATTCGTCAGAGAGATCAGATTTCCCTAATCCGCGACC
QY      781  ASPTYRVALGLULESERILAPROCYSILEPROGLINGLUTHRSERTILELEUPHEAS 800
DB      2341 GATPATGTGAAATATCGGACCTTGATCCCAAAACATCATATCTTTTCGAT
QY      801  LEUSERGLYLSILEVALMETLYSANSERLEUSERIALGLYHISGLYARGMETASPV 820
DB      2401 CTGTAGGCAAGATGTATGATGAAGATAGTTATACGGCGGCATGCGCAATGATGTC
QY      821  SERATGLEUPROASNGILVALATYRILELEULYSVALASGILYRTHRTHRYSILEAS 840
DB      2461 AGCGACACTTCTAATGGGCGCTACATCTTAAGGTGATGATATACGACGAAATTAAT
QY      841  ILEVALHIS 843

```

```

DB      2521 ATAGTCAC 2529
RESULT 2
ID      AAC67909 standard; DNA; 2532 BP.
AC      AAC67909;
DT      16-FEB-2001 (first entry)
XX      Porphyromonas gingivalis periodontal.
DE      Porphyromonas gingivalis; periodontal; antiinflammatory; antibacterial;
KM      amido-lytic; alpha-1-proteinase inhibitor; periodontitis; gingivitis; ds.
OS      Porphyromonas gingivalis.
PN      W0200063394-A2.
PD      26-OCT-2000.
PF      20-APR-2000; 2000MO-US10574.
PR      21-APR-1999; 99US-0130436.
PA      (UYGE-) UNIV GEORGIA RES FOUND INC.
PA      (TRAY/) TRAVIS J.
PA      (POTE/) POTEMPA J.
PA      (NELS/) NELSON D.
PI      Travis J, Potempa J, Nelson D;
DR      WPI: 2000-679600/66.
DR      P-PSDB: AAB36102.
XX      Novel oral bacterial periodontal polypeptide for treating periodontal
PT      diseases, has amidolytic activity for cleavage of non-denatured human
PT      alpha1-proteinase inhibitor at reactive site loop region of inhibitor
PT      -
XX      Claim 24; Fig 5; 55pp; English.
XX      The present sequence is given in a specification relating to novel
CC      oral bacterial polypeptide referred to as periodontin. The polypeptide
CC      has amidolytic activity for cleavage of denatured polypeptides and
CC      non-denatured serpin polypeptides. It has amidolytic activity for
CC      cleavage of a non-denatured human alpha-1-proteinase inhibitor at a
CC      reactive site loop region of the inhibitor. Periodontin is useful for
CC      inhibiting the peptidase activity and reducing periodontitis, loss of
CC      tooth attachment and periodontal pocket formation, and for reducing
CC      growth of bacteria, preferably P. gingivalis in vitro or in vivo.
CC      It is useful for protecting an animal from a disease caused by
CC      P. gingivalis and for treating periodontal diseases, including
CC      gingivitis and periodontitis.
XX      Sequence 2532 BP; 724 A; 561 C; 608 G; 639 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0 Length: 2532
Score: 4438.00 Matches: 843
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-10-030-330-1 (1-843) x AAC67909 (1-2532)
QY      1 MetLysLysSerPheLeuLeuAlaIleValMetLeuPheGlyIleAlaMetGInGlyHis 20
DB      1 ATGAAAAAAGTTTCTTTTACCATGATGATGCTTTTGGCATTTGCCATGACGAGACAT 60
QY      21 SerAlaProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGInVal 40

```

```
Db 61 TCGCTCCGGTACGAAAGCGAGCTTGAGCTGGCTGGCTGGCTTCCGACAGCTA 120
Qy 41 SerLeuArgMetCylGlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArg 60
Db 121 TCCCTTCGGAATGGGAAACACAGCAGTATCTGCACAAAGATTTCATCGATTACGTTTATCGG 180
Qy 61 GlnGlyAspAlaGlyArgGlyIlePheSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 80
Db 181 CAGAGATGCTGAGAGGGGTATCATCATCACAAAGAGAGAGCTCTCCGATTTTAT 240
Qy 81 ValAlaAsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIlePro 100
Db 241 GTAGCTATCTGGGAAATATAGAGGCTATGCTTGTACACACATGACAGAAATACCG 300
Qy 101 ThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAspAsnLeu 120
Db 301 ACAATTTTAGCCATTTCACCCATTTGGCCGTTTGCATGACAGACATGCGGACAACTT 360
Qy 121 ArgMetTyrLeuGlnIleTyrAspGlnGlnIleGlyLeuIleLeuSerGlyValAlaGln 140
Db 361 CGCATGTGCTACAAATTTACGATCAGAAATAGCCGTGATCTTCCGAAAAGCTCAG 420
Qy 141 LeuAsnGlnGlnIleLeuArgThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 420
Db 421 CTGATGAAAGATATATTAGCTACCGAGGCGTACCGGCTGAAAGTACATGCTCTGATGAT 480
Qy 161 AsnGlyHisPheAlaAsnAspProMetArgTyrPheAsnGlnGlyTyrProTyrPheAsnLys 180
Db 481 AACGGTATTTGGCCACAGATCCCATGGGATGAGATCAGATGATACCATGGAACAAATAG 540
Qy 181 GluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyValAlaThrAlaAlaAla 200
Db 541 GAAACACGCTTCTCTAATGGAATCATGCTATACGGCGTGTCTCTACCTGCTGAGCA 600
Qy 201 GlnIleMetArgTyrHisSerTyrProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 600
Db 601 CAATCATGCGCTACCTACCTGCGCGCTTCAAGGATGAAAGGCTTTGCATATATCATCA 660
Qy 221 GlySerLeuValGlyAsnTyrSerGlyIlePheGlyValMetTyrAspTyrIleAsnMet 240
Db 661 GGTTCATTTAGTGGCAACTGCTGCGCACATTTGGTGAATGTATGACATGATCATATG 720
Qy 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAlaAlaTyrAlaThrLeu 260
Db 721 CCGGAAATCCGACCTTGATATCTGACTCAATCTCAAGTGCATGCGTACCGCACACTG 780
Qy 261 MetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
Db 781 ATGCGTGAATGATGATGCTGTTCCATGATGATTTTATGAAATGGAATGATGATGATG 840
Qy 281 SerValTyrValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeu 300
Db 841 AGCGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
Qy 301 HisValAlaGlyAlaLeuTyrThrSerGlnGlnIlePheAspMetIleArgGlyIleLeuAla 320
Db 901 CATGTACGCGCTTATATCTCATCAGAGAGTGCACAGATATGATCCGCGGGAATCTGCC 960
Qy 321 SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCys 340
Db 961 TCCGGAAGGCGGCTATTTATGCAAGGAAATACAGAGATGAGCAATGCTTGTGTGC 1020
Qy 341 AspGlyTyrAlaSerAspLysThrPheHisPheAsnTyrPheGlyValSerAsn 360
Db 1021 GATGCTTATGCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
Qy 361 GlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlnGlnGlnGln 1140
Db 1081 GGCTTCTCAAACTAACACTCTCTCCGCGCACTTCTGTTGGTATCGGAGGTGAGGAAATA 1140
Qy 381 GlyPheThrIleTyrGlnGlnIleIleThrGlyIleGlnProAlaLysThrProAlaGln 400
Db 1141 GGTTTACATTTATCAAGAGATCATCACCAGTATGGAACCGCTTAAAGCTCCCGCTGAA 1200

Qy 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlnAlaGlnIleTyrLysSer 420
Db 1201 GCGGTTACAGATGCTTGGCCGATCTTGGCACTGAAAGACATGAAAGCCAGATGAAAGT 1260
Qy 421 GluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlnGlnGlnGlnGlnGln 440
Db 1261 GAATCCGAGATGAACGATAGGCTATGATATATATACGGGGAAGACCAATCAATCTT 1320
Qy 441 AspLeuGlyTyrArgLeuAsnLysAlaAspGlyValIleGlyValLysThrSerSer 460
Db 1321 GACCTCGATACAGATGAAACAGGCTGACGAGAAAGTACATAGAGGGAACCAATCAATCTT 1380
Qy 461 IleAsnIleSerTyrTyrGlyIleGlyLysIleAspProGlnSerPheSerLeuAlaProAsn 480
Db 1381 ATCAATATCTCTTGGTACGATACGAGAGCATCCGAGAGCTTCTCATTTGGCACCTAAT 1440
Qy 481 GlnLeuSerGlnGlyIleAsnThrIlePheLeuLeuTyrArgArgThrGlyThrGlnGln 500
Db 1441 CAGTTGTCAAGGAATCAACACCATCACTTATCTGTCGACAGCACCGCAACG 1500
Qy 501 TrrGluProValArgHisAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
Db 1501 TGGGAGCGGCTACGCGCATGCGACAGGAGATATGCAATAGCATTTAAAGTAAATACGCA 1560
Qy 521 AspProAsnAsnValValIleThrValAspAsnGlnGlyLysLeuSerIleValPro 540
Db 1561 GACCCGAAACAATGCTAGTACAGTAAATTAACGAAGGCAACCTCATGATGCTCCC 1620
Qy 541 AsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPheAsn 560
Db 1621 AACAGCTTGTGCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
Qy 561 SerAspSerProAspGlnIleArgThrProValAlaPheAlaLeuSerThrGlyAlaThr 580
Db 1681 AGGACAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740
Qy 581 AlaAspAspValIleSerLeuGlyTyrValMetAlaGlnValProGlyLysSerSerAsn 600
Db 1741 GCGAGCATGATGATATCTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1800
Qy 601 TyrProValIleTyrSerLysAspValLeuThrLeuSerGlnGlyAspTyrThrLeuThr 620
Db 1801 TATCCGAGGTTTGGTGTAAAGAGCTTCTCACTCTCCGAAAGCGAGCATATCATTTGG 1860
Qy 621 TyrArgPheSerIleAsnAsnGlnLysAspGlnTyrLysIleGlySerValSerVal 640
Db 1861 TATGATTTTCTCAATCAACCAAAAGATGATGAAAGAAAGATGGAAGGCTGTCAGTA 1920
Qy 641 LysThrProThrGlnTyrThrHisProLeuPheGlnValGlyHisAsnGlnThrSerThr 660
Db 1921 AAAACACCGAGATATGCAACCCCTTATTCGAAAGGGCCCAATATCAAACTTACAC 1980
Qy 661 TyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeuGly 680
Db 1981 TATACCTGATATGATGACACACAGATATTTGCCGACTTATCACTCAAAAATCTCGGA 2040
Qy 681 LeuProPheAsnGlyGlnLeuValValPheArgGlnThrGlnSerSerSerGlySer 700
Db 2041 TTGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100
Qy 701 LeuThrAlaAlaGlnGlnIleThrValHisIleLysGlnGlnGlnGlnGlnGlnGlnGlnGln 2160
Db 2101 TTATGGGAGCTCAAGAAACAGATACATCAACAGCAAGAAACTTGTGATATTAACCT 2220
Qy 721 ValValGlnGlnProIleProAspGlySerTyrArgAlaThrLeuHisAlaPheValAsn 740
Db 2261 GTTGTGAAAGCCCTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280
Qy 741 GlyGlnGlnGlnLeuTyrLeuLysGlySerArgAsnTyrThrValLysIleValAsnGly 760
Db 2221 GGACAAACAAGTATGATCTCAAGGGAAGAAAGTACAGCGTGAAGATCTCATATGATG 2280
```

QY 761 ThrAlaValGluAlaIleGluSerSerGluIleArgValPheProAsnProAlaArg 780
DB 2281 ACAGGGGTAGAGCAATGAAATCGACAGAGATCAGATTTCCCTAATCCGGCAGCC 2340
QY 781 AspyrValGluIleSerAlaProCysIleProGluIleThrSerIleIleuPheAsp 800
DB 2341 GATTATGCGAAATCGGCACCTTGCAATCCCAAGAACATCTATCTTTTCGAT 2400
QY 801 LeuSerGlyLysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspVal 820
DB 2401 CTGTCAGGCAAGATTCATGATGAGATGTTATCAGCGGGGCGCATGCGCAATGGATGTC 2460
QY 821 SerArgLeuProAsnGlyAlaTyrIleLeuLysValAspGlyTyrThrLysIleAsn 840
DB 2461 AGCGGACTTCCTAATGGGCGCTACATCTTAAGTGATGATATACGAGAAATTAAT 2520
QY 841 IleValHis 843
DB 2521 ATAGTGCAC 2529
DB 1578
DB AAX91578 standard; DNA; 2547 BP.
AC AAX91578;
AT 25-AUG-1999 (first entry)
DE Porphyromonas gingivalis protein PG28 ORF encoding DNA.
XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic; ds.
XX Porphyromonas gingivalis.
OS Porphyromonas gingivalis.
PN W09929870-A1.
PD 17-JUN-1999.
XX 10-DEC-1998; 98WO-AU01023.
PF 10-DEC-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX
XX (CSLIC-) CSL LTD.
PA Aglus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX MPI: 1999-385613/32.
DR P-PSDB; AAY34360.
XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
PS Claim 12; Page 124; 588bp; English.
XX AAY91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC assays. Porphyromonas gingivalis is involved in periodontal disease

CC especially gingivitis.
XX
SQ Sequence 2547 BP; 727 A; 566 C; 610 G; 644 T; 0 other;
Alignment Scores:
Pred. No.: 0 Length: 2547
Score: 4438.00 Matches: 843
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 20 Gaps: 0
US-10-030-330-1 (1-843) x AAX91578 (1-2547)
QY 1 MetLysLysSerPheLeuLeuAlaIleValMetLeuPheGlyIleAlaMetGlnLysHis 20
DB 19 ATGAAAAAAGTTTCTTTTACCAATGTAATGCTTTTGGCAATTCCTCATGCGGACAT 78
QY 21 SerAlaProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnVal 40
DB 79 TCTGCTCCGGTTACGAAAGAGGAGCTTGAGCTGCTGCGCTGCTTGGCAGAGTA 138
QY 41 SerLeuArgMetGlyGlnThrAlaValSerAspLysIleSerIleAspyrValTyrArg 60
DB 139 TCCTTGCGAATGGGCAAAACAGCAGATCTGCACAGATTTCCATTCATTCGTTATCGG 198
QY 61 GlnGlyAspAlaGluArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyr 80
DB 199 CAAGGAGATCGAGAGGGGATACATCATCACAAGAGGCTCTCTCATATTTTAT 258
QY 81 ValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuValAlaAlaAspArgIlePro 100
DB 259 GTAGCTATCTGGAATATATAGAGGCTTGTCTTGACAGACATGACAGAAATACCG 318
QY 101 ThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAspAsnLeu 120
DB 319 ACAATTTTAGCCTATTCACCCATTTGGCCCTTTCGACATGAGATGCGGACAAATCTT 378
QY 121 ArgMetTrpLeuGlnIleTyrAspGlnGluIleGlyLeuIleLeuSerGlyLysAlaGln 140
DB 379 CGCATGTGGCTCAAAATTTAGCATGAGAAATAGGCTCATATCTTCCGAAAGCTCAG 438
QY 141 LeuAsnGluGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMetAsp 160
DB 439 CTCATGAAAGATATATACCTACCGAGGCGTCACCGCTGAGTACATGCTCTGATGAT 498
QY 161 AsnGlyHisPheAlaAsnAspProMetArgTyrPasnGlnGlyTyrProTyrPasnAsnLys 180
DB 499 AACGGTCATTTTGGCAACGATCCCATGCGATGGAATCAAGTTACCATGGAACATTAAG 558
QY 181 GluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAlaThrAlaAlaAla 200
DB 559 GAACCACTGCTCTAATGCAATCAGCTTACCGGCTGTCTCTCTCTCTCTCTCTCTCT 618
QY 201 GlnIleMetArgTyrHisSerTrpProLeuGlnGlyGlySerPheAspyrHisAla 220
DB 619 CAATATCATCGCTACCATATGCTGCGCTTCAAGGTGAAGGCTCTTTCATTTATCTATGCA 678
QY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsnMet 240
DB 679 GGTTCATTTAGTTGGCAACGTGTCGGCCATTTGGTAATGACATGATGATATG 738
QY 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThrLeu 260
DB 739 CCGGGAATCCGACCTTATATCTGACTCAATCAAGTGAATGATGATGATGATGATG 798
QY 261 MetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
DB 799 ATGCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 858
QY 281 SerValTyrValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeu 300
DB 859 ACGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 918

OY 301 HisValargAlaLeuTyrThrSerGlnGluTrpHisAspMetCIIArgGlyGluLeuAla 320
 DB |||||
 OY 919 CATATACGGCCCTTAAATACCTCACAGAGTGGCCACGATATCATCCGGGGAAACTTGCC 978
 DB |||||
 OY 321 SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCys 340
 DB |||||
 OY 979 TCGGAAGCCCGGTATATATGACAGGAAATAACAGACATAGACATAGCATGCTTCCTTGC 1038
 DB |||||
 OY 341 AspGlyTyrAlaSerAspGlyThrPheHisAspSerProGlyTyrPglYglValSerAsn 360
 DB |||||
 OY 1039 GATGGTATGCTTGGATGGTACTTCCATTTCACTGAGGCTTGGGGAGGTGTTTCCAAC 1098
 DB |||||
 OY 361 GlyPheTyrLysLeuThrLeuLeuSerProThrSerIleGlyGlyGluGlyIle 380
 DB |||||
 OY 1099 GACTTCTACAAACATCAACACTCTCTCCGACACTCTGTTGGGTATCGGAGCTGAGGAAATA 1158
 DB |||||
 OY 381 GlyPheThrIleTyrGlnGluIleIleThrGlyIleGluProAlaLysThrProAlaGlu 400
 DB |||||
 OY 1159 GGTATTACATTTATCAAGAGATCATCCCGATATCGAACCGGCTTAACTCCGCTGAA 1218
 DB |||||
 OY 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluAlaGluTyrLysSer 420
 DB |||||
 OY 1219 GCCGTATCAATGATCTCTGGCATCTGGCATGAAAGCATAGAAAGCCGAGTATAAAGT 1278
 DB |||||
 OY 421 GluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGluGluGlnSerAsnLeu 440
 DB |||||
 OY 1279 GAATTCGGATGTACGTAGGATTCGATATATATACAGGTGAAAGACATCAATCAATCTT 1338
 DB |||||
 OY 441 AspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluValIleGluValLysThrSerSer 460
 DB |||||
 OY 1339 GACCTCGGATACAGATTATACACAGGCTGACGAGAAATCATAGAGTGAAACTTCATCT 1398
 DB |||||
 OY 461 IleAsnIleSerTyrTyrGlyTyrGlyLysIleProGluSerPheSerLeuAlaProAsn 480
 DB |||||
 OY 1399 ATCATATCTCTGTGATGAGTACGAGACATCCCGAAGTTCATTTGCTGACCTAT 1458
 DB |||||
 OY 481 GlnLeuSerGlnGlyIleAsnThrIleThrLeuTyrArgArgThrGlyThrGlnGln 500
 DB |||||
 OY 1459 CAGTTGTACAGAAATCAACACCATCCCTTACTTATCTGTCACAGGACCCGAAACG 1518
 DB |||||
 OY 501 TrpGluProValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsnThrThr 520
 DB |||||
 OY 1519 TCGGAGCGGCTACGATGACAGAGAGATATGTCATATACATTAAAGTAAATACACA 1578
 DB |||||
 OY 521 AspProAsnAsnValValIleThrValAspAsnGlnGlyLysLeuSerIleValPro 540
 DB |||||
 OY 1579 GACCCGAAACAATGTCGTAGTACGCTATATTAACAGAGCAAGCTATATGTCCTCC 1638
 DB |||||
 OY 541 AsnSerPheValAlaAspLeuAsnSerTyrGluHisSerThrIleThrValGlnPheAsn 560
 DB |||||
 OY 1639 AACACCTTGTGCGAGATCTGAATCTTATGAACATGTGATGATTCACATGCTTCAT 1698
 DB |||||
 OY 561 SerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGlyAlaThr 580
 DB |||||
 OY 1699 AGCCAGACCCCTGATGTAGATCCGTACACCCGTAGCTTGTCTCTACAGAGACTTACT 1758
 DB |||||
 OY 581 AlaAspAspValIleSerLeuGlyTyrPheAlaGluValProGlyLysSerSerAsn 600
 DB |||||
 OY 1759 GCGGACGATGATATCTTGGCTGGGTAAATGCTGAAGTTCCGGCGGCTAGACCAAC 1818
 DB |||||
 OY 601 TyrProValValTyrSerLysAspValLeuThrLeuSerGluGlyAspTyrThrLeuTyr 620
 DB |||||
 OY 1819 TATCCGCTGTTGGTTGAAGAGCTTCTCCTCCGAAAGCGACTATATGTTG 1878
 DB |||||
 OY 621 TyrArgPheSerIleAsnAsnGlnLysAspGluTyrLysLysIleGlySerValSerVal 640
 DB |||||
 OY 1879 TATAGATTTTCCATCAACACCAAGAGATGAGAAAAAGATCGAAGCCGTCTAGTA 1938
 DB |||||
 OY 641 LysThrProThrGlyTyrThrHisProLeuPheGluValGlyHisAsnGlnThrSerThr 660
 DB |||||
 OY 1939 AAAACACCGACAGATACGACACCCCTTATGGAAGTGGCCATATATCAAACTTCTAC 1998
 DB |||||

OY 661 TyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeuGly 680
 DB |||||
 OY 1999 TATACCGGATATGAGACACAAACAGATTTCCCGACCTTACACTCAAAAATCTCCGA 2058
 DB |||||
 OY 681 LeuProPheAsnGlyGluLeuValValPheArgGluThrGlnSerSerGlySer 700
 DB |||||
 OY 2059 TTGGCTTTCAATGGTGTGTTGGTTGTTTCCGCCAAACATCTCATCGGGGCTT 2118
 DB |||||
 OY 701 LeuTrpAlaAlaGlnGluThrValHisIleLysGlnGlyGluThrPheValTyrLysPro 720
 DB |||||
 OY 2119 TTAGGGGACCTTAAAGAAACAGTATATCAAGCAAGAGAAACTTTCGTATTAACCT 2178
 DB |||||
 OY 721 ValValGluGlyProIleProAspGlySerTyrArgAlaThrLeuHisAlaPheValAsn 740
 DB |||||
 OY 2179 GTTGTGAAAGGCCCTATACCTGATGATCCATCTGTCGACCCCTCATTCATTCGTAAAC 2238
 DB |||||
 OY 741 GlyGlnGlnGlnLeuTyrLeuLysGlyLysArgAsnTyrThrValLysIleValAsnGly 760
 DB |||||
 OY 2239 GGACAAACAAGTTGTACTTCAAGGGGAAAGAACTACACGGTGAAGATCTCATATGCT 2298
 DB |||||
 OY 761 ThrAlaValAlaGluAlaIleGluSerSerGluGluIleArgValPheProAsnProAlaArg 780
 DB |||||
 OY 2299 ACACGGGTAGAGCAATAGATGCTCAGAAAGATCAGAGTATCCCTTAATCCGGCACGC 2358
 DB |||||
 OY 781 AspTyrValGluIleSerAlaProCysIleProGlnGluThrSerIleIleLeuPheAsp 800
 DB |||||
 OY 2359 GATTATGTGAAATATCGGCACCTTCATTCCTCCCAAGAAACATCATCATCTTTCGAT 2418
 DB |||||
 OY 801 LeuSerGlyLysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspVal 820
 DB |||||
 OY 2419 CTGTACAGCAAGATTTCTATGAAAGATAGTTATACGGGGCATGACCAATGGATGTC 2478
 DB |||||
 OY 821 SerArgLeuProAsnGlyAlaTyrIleLeuLysValAspGlyTyrThrLysIleAsn 840
 DB |||||
 OY 2479 AGCCGACTTCCTTAAGGGGCTTACATCTTAAAGTGTGATGATATACGAAATAAAT 2538
 DB |||||
 OY 841 IleValHis 843
 DB |||||
 OY 2539 ATAGTGAC 2547
 DB |||||

RESULT 4
 ABN70210
 ID ABN70210 standard; DNA; 1194 BP.
 XX
 AC ABN70210;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Streptococcus polynucleotide SEQ ID NO 8333.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 OS antinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX
 PN Streptococcus pyogenes.
 XX
 MO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001MO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 XX (GENO-) INST GENOMIC RES.
 XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettelin H;
 DR WPI; 2002-352536/38.

PF 30-APR-1996; 96WO-US05997.
 XX
 PR 01-MAY-1995; 95US-0432692.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Ananthaswamy HN, Fernandez A, Kapur V, Nusser JM;
 XX
 DR WPI: 1996-506148/50.
 DR P-PSDB; AAM07898.
 XX
 PT Use of extracellular Streptococcal cysteine protease enzyme - for
 PT inhibiting the proliferation of neoplastic cells, e.g. for treating
 PT carcinoma, lymphoma or leukemia.
 XX
 PS Disclosure: Pages 62-63; 99pp; English.

The present sequence encodes the Streptococcus pyogenes clone speB7
 pre-pro cysteine protease (CP), which can be used to inhibit
 neoplastic cell proliferation, especially in a human, useful in the
 treatment of neoplastic conditions, e.g. carcinomas, sarcomas,
 melanomas, lymphomas and leukemias originating from blood, lung,
 mammary gland, prostate, intestine, stomach, liver, heart, skin,
 pancreas or brain tissue. The CP is especially associated with a
 wound covering, and can also be used to prevent metastasis or
 identify susceptible neoplastic cells.
 CC K1735 and CM519 melanoma cells were injected s.c. into nu/nu mice,
 CC optionally followed by i.p. injection of CP (100 microg, 24 hours
 CC later). The mice were checked twice weekly for tumor growth for
 CC 12 weeks, to give results that showed that treatment with CP
 CC completely protected athymic mice against transplanted K1735
 CC melanoma growth, and protected 60% of the mice from developing
 CC CM519 melanoma.

XX Sequence 1197 BP; 393 A; 239 C; 235 G; 330 T; 0 other:

Alignment Scores:

Pred. No.:	1-796-31	Length:	1197
Percent Similarity:	484.00	Matches:	125
Best Local Similarity:	31.49%	Conservative:	72
Query Match:	10.91%	Mismatches:	138
		Gaps:	63
			13

US-10-030-330-1 (1-843) x AAT5219 (1-1197)

```

26  LysGluAArgAlaLeuSerLeuAlaArgLeuAlaLeuAlaValSerLeuArgMetGly 45
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107  AAGAAGCAAAAGATGCGCTATACATTATCCAAAATACAGCA-GCTATCAAAACAGGT 165
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46  GlnThrAlaValSerAlaPheSerLeuAspTyrValTyrArgGlnGluAlaGlu 65
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
166  GCACGACGCCGCAAGAT---ATTAGCTTGCAAAAGTTAACTTAGTGGAAGACT--- 219
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
66  ArgGlyIleThrSerGlnGlnGluSerProAlaTyrPheTyrValAlaAsnArgGly 85
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
220  -----TCTGCTCTAATATGTAATGTTTACAAATATTCTACTGGA 258
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
86  AsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
239  -----GGAATTTGTTATCGTTTCAGAGATAAAGCTTCCGAAATCTTAGATAC 309
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
106  SerProIleGlyArgPheAspMetAspSerMetProAspAlaLeuArgMetIlePheGln 125
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
310  TCTACACACGATCATTTTGACGCTTAACGT---AAAGAAACATTTGCTTCTTACGAA 366
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126  IleTyr-----AspGlnGluIleGlyLeuIleLeuSerGlyTyrAla 139
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
367  AGTATGTGCAACAATCAAGAAACAAATAATTAGACACTTATGCTGTCGTCGCT 426
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
140  GlnLeuAsnGlnGluIleLeuArgThrGlnGlyValProAlaGluValIleAlaLeuMet 159
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

427  GAGATTAAACACAGTT-----GTTAAATCTCTCTCT 459
160  AspAsnGlnIleHisPheAlaAsnAspProMetIleArgTrpAsnGlnGlyTyrProTrpAsn 179
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
460  GAT-----TCAAAAGCATTCATTACACACAGATTAACCTTACACCTA 504
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
180  LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
505  TTGACACCTGTTATTGAAAGGTAACACAGGTGAAATCTTTTGTAGTCACATGCA 564
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
191  TyrThrGlyCysValAlaIleThrAlaAlaIleGlnIleMetAlaGlyTyrHisSerTrpLeu 210
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
565  GCTACAGGATGTGTGCTACTGCAACCTCTCAATATGAATATCAATTAACCT--- 621
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
211  GlnGlyGlnGlySerPheAspTyr----- 218
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
622  ---AACAAAGGTTGAAAGACTACCTTAGACACTTAACCTCAATTAACCATTTTCAAC 678
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
219  HisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGlnGluMetTyrAspTrpIle 238
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
679  CATCTTAAGAACTTG-----TTTGACAGTATCTCTACTAGACAAATACATGCAAC 729
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
239  Asn---MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyr 257
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
730  AACATCTTACTTACTTATAGCGGAAGAAATCTTAAGCTTCAAAAATG-----GCGAT 783
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
258  AlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTyrGlnIleAsnGlySer 277
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
784  TCAGAAATGATGCGTATGTTGTTGTTTTCAGTACAGACATGATATGCTCATCTGTCGT 843
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
278  GlyThrTyrSerValTyrValAlaGlyAlaLeuArgAsnAsnPheArgTyrIleArgSer 297
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
844  TCTGCAGGTAGCTCTCGTGTCAAGAGCCTTGAAGAAACTTTGGCTACAAACCAATCT 903
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
298  LeuGlnLeuHisValArgAlaLeuTyrThrSerGlnGluTyrPheHisAspMetIleArgGly 317
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
904  GTTCACCAAAATTAACCGTACGACTTATGCAAAAGATGGAGACCAAAATTAACAA 963
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
318  GluLeuAlaSerGlyArgProValTyrTyrAlaGlnIleAsnGlnSerIleGlyHisAla 337
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
964  GAAATATCTCAAAACCAACAGATATACCAAGATGTCGGAATAGTGGCGACATGCC 1023
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
338  PheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrIleGly 357
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1024  TTTGTTATGCAAGTGTGTCAGCGAGCTAACTTACATGTTAACTGGGCTGGGTGGA 1083
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
358  ValSerAsnGlyPheTyrIleThrLeuLeuSerProThrSerLeuGlyIleGly 377
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1084  GTCTGTGACGGCTTCTTCGTTTGACGCACTAAACCTTCAGCTTGTGTAAGTCTGTCG 1143
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
378  GlnGlyIleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1144  GCGCAGCGCGCTTCAACGCTTACCAAGTCTGTGTGATGACATCAAACT 1194
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

RESULT 6

AAT15294
 ID AAT15294 standard; cDNA; 1197 BP.

AC AAT15294;
 XX 30-OCT-1996 (first entry)

DE S. pyogenes speB extracellular protease gene.

XX Immunogenic peptide: speB gene; extracellular protease; production;
 XX antibody; vaccine; diagnosis; detection; Streptococcus infection;
 KW group A; prevention; treatment; pharyngitis; tonsillitis;
 KW skin infection; acute rheumatic fever; scarlet fever; probe;
 KW post-streptococcal glomerulonephritis; sepsis; meningitis;
 KW erysipelas; cellulitis; fasciitis; toxic shock like syndrome; ss.
 XX Streptococcus pyogenes.

D	b	367	AGTATATGTCGACCAACATCAAGAAAGAAACAAAATTTAGCACTACTTATAGCGTGACCGCT	426
Q	y	140	GlnduabnsglgluIuileuargThrgluIyValProalaglValhIsalaleuMet	159
D	b	427	GAGATTAAACACCAAGTT-----GTTAAATTCCTCCTT	459
Q	y	160	AspsnsglYhIsPhehlaAsnAspPrometLgrtTrrpanGlnglyTyrrProTrrpanSn	179
D	b	460	GAT-----TCMAAAGGCATTCATTACACCAAGGTAAACCCCTTACAACTTA	504
Q	y	180	LysgluProlLeuLeuProasn-----GlyAsnHsAla	190
D	b	505	TTGACACCTGTTATTGAAAAAGTAAACACAGGTACACATCTTTTGTAAGTCACATGCA	564
Q	y	191	TyrThrgIcYcValAlaIaThAlaAlaIaGlnIleMetargTyrrHsSerTrrProLeu	210
D	b	565	GCTACAGAGTGTGTGTCTACTGCAACAGCTCAAAATTATGAATATCATAAATTACCTT	621
Q	y	211	GlnglygluIySerPheAspTyrr-----	218
D	b	622	--AACCAAGGGTTGAAAGACTTACACTTGAACACTTAAGCTCAAAATTAACCATATTTCAC	678
Q	y	219	HIsalaglSerleuValglYAsnTrrpSerIyThrPheglYgluMetTyrrAspTrrle	238
D	b	679	CATCCCTAAGAACTTG-----TTTGACACTTCTCTACTATGACATACAACTGGAAC	729
Q	y	239	Asn--MetProglYasnProAspLeuAspAsnleuthrGlnSerGlnValAspAlaTyrr	257
D	b	730	AACATCTCACTACTTATTAGCGGAAGAAAGAACTTAACGTTCAAAAATG-----GCGATT	783
Q	y	258	AlaThrleuMetLargAspValSerAlaSerValSerMetSerPheTyrrGluAsnGlySer	277
D	b	784	TCAGAAATTGAGGGCTGATGGTGGTATTTCACTAGACATGATGATATGGTCATCTAGTGGT	843
Q	y	278	GlyThrTyrrSerValTyrrValIaIglYAlaLeuArgAsnAsnPheargTyrrLysArgSer	297
D	b	844	TCCTGACAGTACCTCTCGTTCACAAAGACCTTGAAAGAAACTTGGCTACAAACCAATCT	903
Q	y	298	LeuGlnleuHIsValArgAlaLeuTyrrThrSerGlnglnIuTrrHIsAerMetIleArgly	317
D	b	904	GTTTCACCAAAATTAACCCGTAGCGACTTTAGCAACAAAGATTTGGGAAGCAACAAATTCACAAA	963
Q	y	318	GluIeuAlaSerGlyArgProValTyrrTyrrAlaIaLeuArgAsnAsnInserIleGlyHIsAla	337
D	b	964	GAATATTCACAAACCAACGATACTACCAAGGTGCGGTAAAGTAGGCGGACATGCC	1023
Q	y	338	PheValCysAspGlyTyrrAlaSerAspIyThrPheHIsPheAsnTrrPglYtTrrPglYglY	357
D	b	1024	TTTGTATTGCAATGAGTGGTCGACGAGCACTTACATCAACATGTAAGTGGGGTGGGATGCA	1083
Q	y	358	ValSerAsnGlyPheTyrrLysLeuThrLeuLeuSerProthrSerleuGlyIleGlyGly	377
D	b	1084	GTCCTTCACGCGCTTCCGCTTCAGCCACCAATCAACCCCTGACGCTTGTGATACGTTGGCG	1143
Q	y	378	GluIuIyIleGlyPheThrIleTyrrGlnglnIuIleIleThrglyIleGluPro	394
D	b	1144	GGCGACAGCGGCTTCACGCTTACCAAGTCTGTTAGGCACTAAACCT	1194
RESULT 7				
AAAO7111				
ID	AAAO7111 standard; cDNA; 1197 bp.			
XX	AAAO7111;			
XX	09-JUN-2000 (first entry)			
DE	S. pyogenes cysteine protease speB7 coding sequence.			
XX				
KM	Cysteine protease; speB; Group A Streptococcus; extracellular protease;			
KM	detection; diagnosis; extracellular matrix; infection; skin infection;			
KM	disease status monitoring; vaccine; Streptococcus mediated disease;			
KM	pharyngitis; tonsillitis; scarlet fever; sepsis; erysipelas; fasciitis;			
KM	pneumonia; acute rheumatic fever; poststreptococcal glomerulonephritis;			

Dd	34	GAGGGCAAGGCACAGGCCGATAGATCCTATCCGTGTGAAGCGAAGCCCTACACAGGGAGACTGGCA	288
Qy	153	AlaGluValAlaHisAlaLeuMetCaspAsnGlyHisIlePheAlaAsnAspPromeTatGTPAsn	172
Dd	287	TCATTCATTCATCCCTCTTTTGGGAACGGCGCAACATGCATCGAGATCCGATCTGTGGGAT	228
Qy	173	GlnGlyTyrProTPrAsnAsnGlySerGluProLeuIleProAsnGlyAsnHisAlaTyrThr	197
Dd	227	CAGGGCATTCATTTATTAACACTTTCGATCCCTCCGTCTCTCCGGGACAGGCTTATACC	166
Qy	193	GlyCysValAlaIleThrAlaAlaAlaGlnIleMetArgTyrHisSerTPrProLeuGlnGly	212
Dd	167	GGTTGTGTGTCCACTCCATCGGACAAATCATCGCTCACTACAAATGGCCGAGAGAGCT	108
Qy	213	GluGlySerPheAspTyrHisAlaGlySerLeuValGly-----AsnTPrSerGly	225
Dd	107	TCCGGCGGAATATGACTACTAC--GATGATATGACGGGAGCAGCATACCCTATTCAGGC	51
Qy	230	ThrPheGlyGlnMetTyrAspTPrIleAsnMetProGlyAsn	243
Dd	50	ACATTCGGTGCAGCACTATACGTGAGCAAGATGCCCGGTAC	9
RESULT	9		
AAAX91791			
ID	AA911791 standard; DNA; 1620 bp.		
XX	AA911791;		
AC			
XX	25-AUG-1999 (first entry)		
DT			
XX	Porphyromonas gingivalis protein PG91 encoding DNA.		
DE			
XX	Porphyromonas gingivalis; Pg; periodontal disease; gingivitis;		
KW	Porphyromonas gingivalis; ds.		
KW			
XX	Porphyromonas gingivalis.		
OS			
XX	MO9929870-A1.		
PN			
XX	17-JUN-1999.		
PD			
XX	10-DEC-1998; 98MO-AU01023.		
PF			
XX	04-AUG-1998; 98AU-0005028.		
PR	10-DEC-1997; 97AU-0000839.		
PR	31-DEC-1997; 97AU-0001182.		
PR	30-JAN-1998; 98AU-0001546.		
PR	10-MAR-1998; 98AU-0002264.		
PR	09-APR-1998; 98AU-0002911.		
PR	23-APR-1998; 98AU-0003128.		
PR	05-MAY-1998; 98AU-0003338.		
PR	22-MAY-1998; 98AU-0003654.		
PR	29-JUL-1998; 98AU-0004917.		
XX			
PA	(CSLC-) CSL LTD.		
XX			
PI	Agus CT, Barr IG, Hocking DM, Margetis MB, Patterson MA;		
PI	Ross BC, Rothel LJ, Webb EA;		
XX			
DR	WPI: 1999-385613/32.		
DR	P-PSDB: AAY34573.		
XX			
XX	Antigenic Porphyromonas gingivalis peptides for preventing		
PT	gingivitis		
XX			
PS	Claim 12; Page 270; 588bp; English.		
CC	AA91536 to AA91801 encode two hundred and sixty six antigenic		
CC	Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to		
CC	AAY34583. AA91802 to AA91989 represent PCR primers used in the		
CC	isolation of the Pg polypeptides. The Pg polypeptides have antibacterial		
CC	activity with a vaccine mechanism of action. The Pg polypeptides can be		
CC	used as vaccines especially against Porphyromonas gingivalis. Probes can		

PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX (CSLC-) CSL LTD.
XX
XX Aglus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
XX WPI: 1999-385613/32.
DR P-PSDB; AAY34430.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX
XX Claim 12, Page 172; 588bp; English.

AAK91536 to AAK91801 encode two hundred and sixty six antigenic
Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
AAY34553. AAK91802 to AAK91989 represent PCR primers used in the
isolation of the PG polypeptides. The PG polypeptides have antibacterial
activity with a vaccine mechanism of action. The PG polypeptides can be
used as vaccines especially against Porphyromonas gingivalis. Probes can
be used to detect Porphyromonas gingivalis in standard hybridisation
assays. Porphyromonas gingivalis is involved in periodontal disease
especially gingivitis.

XX SQ Sequence 1689 BP; 505 A; 361 C; 423 G; 400 T; 0 other;

Alignment Scores:

Pred. No.: 2,04e-06 Length: 1689
Score: 190.00 Matches: 103
Percent Similarity: 35.98% Conservative: 78
Best Local Similarity: 20.48% Mismatches: 162
Query Match: 4.28% Indels: 160
DB: 20 Gaps: 23

US-10-030-330-1 (1-843) x AAK91648 (1-1689)

QY 413 AspliegialaglytyrlysergluSerlyleuAsnVal----- 426
DB 370 GATATCGGTAAACGGTAAAGATGATCCGACATCAAGATTGGTCGTAGACGTATG 429
QY 427 -----GlyTyr-----SerIleTyrAsnThrGlygluGlnSer 438
DB 430 AATAAGCCCGAGGATATAGAGTAGAGTGGCGGCTTCAGTCGCGATGCCAAGCGCGAG 489
QY 439 AsnleuAspLeuGlyTyrArgleuAsnlysaAspGlyluValIleGluValylsThr 458
DB 490 AATGGGAACCTGCTATAGAGACCTTCTCAATGTCAGATTGACATGTGATATA 549
QY 459 SerSerIleAsnIleSerTyrGlyTyrGlyGlnHisProGluSerPheSerleuAla 478
DB 550 GCCTTC-----AACTAT-----CGTTCG 567
QY 479 ProAsnGlnleuSerGlnGlyIleAsnThrIleThrleuLeuTyrArgArgThrGlyThr 498
DB 568 CCTTCTCTCTTAACAAATGTCGCAACCTTTCCTTGGCTTTCGCTTACCGGCTTC 627
QY 499 GluGlnTyrGluProValArgHislaGlnGlyTyrValAsnSerIleValAsn 518
DB 627 ----- 627
QY 519 ThrThrAspProAsnAsnValValThrValAspAsnAsnGlylyLysleuSerIle 538
DB 628 -----AACAAATACGACAAATAAATTAAGTTT 651
QY 539 ValProAsnSerPheValAlaAspLeuAsnSerTyrGlyGlnHisSerThrIleThrValGln 558
DB 652 GTG-----GACTATGTGTCTCTCTGAATGAGGCGCAAAATTCATAAAACTTACTC 705

QY 559 pheAsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGly 578
DB 706 TTCAGTCAGATGAGAGAGAAATTTGAC-----AAGTGGATCTCTCATGGGT 756
QY 579 AlaThrAlaAspAspValIleSerleuGlyTyrValMetAlaGluVal----- 594
DB 757 AGCAGCTGTGAATCCATGGGTGCACATGCTGCGCCCTAATGGGTGGTATCGAAATG 816
QY 595 -----ProGlyGlySerSerAsnTyrProValAl----- 604
DB 817 AATAAAGAGGGGAAAAAGCGATATCGGTTCTGTGCAACTTGTGCAAAATGATCC 876
QY 605 -----TyrSer-----Lys 607
DB 877 GAATTTCAGTGGTCAGAGCCCTATAAAGTAGTGAAGCAGCATGTCGTTACGCCCAA 936
QY 608 AspValleuThrleuSerGlnGlyAspTyrThrleu-----TyrTyrArgPhe 623
DB 937 ATCCAAATGTGTGTCGAGCAGATTAACAATACGATCAATGGGAGATTGCCAACCTTC 996
QY 624 SerIle-----AsnAsnGlnLysAspGluTyrLysIleGlySerVal 638
DB 997 ATGATTACGTACAGCGATTATGATCTGATATCCCAAGAAATCTTCAGATGAAAAAGA 1038
QY 639 SerVallyThrProThrGluTyrThrHisProleuPheGluValGlyHisAsnGlnThr 658
DB 1039 -----GACTTCGGGTATGTATATCCCAAGAAATCTTCAGATGAAAAAGA 1086
QY 659 SerThrTyrThrleuAspMetAlaHisAsnArgValleuProAspPhe-----Thrleu 676
DB 1087 AAACCTCCGATATGATGATCTGTGGAAGCTTCTTCACGCTTCGTACAGAGTAG 1146
QY 677 LysAsnleuGlyleuProPhe-----AsnGlyGlnleuValValAlaPheArg 692
DB 1147 ACCAAATCGGGCTGGGTGATGACAGAAAGCCCAATCACTGATGATATGTCGCA 1206
QY 693 GlnThrGlnSerSerSerGlySerleu-----TyrAlaIleGlnGlnThrValHis 709
DB 1207 AAAGAGAGACGGTACGAACAGCTGAATATCCGCTGGCCAAATATGCAAGATTCAT 1266
QY 710 IleLysGln-----GlyGluThrPhe-----ValTyrLys 719
DB 1267 AACAAAGATTGGAGGACGACACATTTAGCTATACATCTGCCAATGCTCTCTACACA 1326
QY 720 ProValValGlnGlyProIleProAspGlySerTyrArgAlaThrleuHisAlaPheVal 739
DB 1327 CCTCAAGTAGAC-----ATC 1341
QY 740 AsnGlyGlnGlnleu-----TyrleuLysGlyLysArgAsn 752
DB 1342 AATCCGACCAAGGCTCTGTCGTCGTCATGGGTGGAATATCTGCCCGCAACCGATC 1401
QY 753 Tyr-----ThrVallyIleValAsnGlyThrAlaValGluAlaIleGluSerGly 770
DB 1402 GTTGGTGTATACGACGATGACCATGCCAAGCGGTAGAGACATCGTAATGCAAGAA 1461
QY 771 -----GluIleArgValPheProAsnProAlaArgAspTyrValGluIleSerAlaProCys 789
DB 1462 GGCAGCATACCTTACCCGAAATTCGCTCAAGAAATATGCTGTATACCTCGCGC 1518
QY 790 IleProGlnGluThrSerIleIleleuPheAspLeuSerGlyLysIleValMetLysAsn 809
DB 1519 ACGGACGAAACTGCAAGGCTGTGTATTACGATATGACAGGCGAGATAGTCGTAGGCT 1578
QY 810 SerleuSerAlaGlyHisGlyArgMetAspValSerArgleuProAsnGlyAlaTyrIle 829
DB 1579 TCTTCTCCGCAAGCAATACAGGCTGACAGTCACCACTTGTGCTAAGGATACGTACATA 1638
QY 830 leuLysVal 832
DB 1639 CTCAGGTC 1647

RESULT 11

Oy	363	yrLysLeuThrLeuLeuSerProThiSerLeuGlyIleGlyGly	-----	377
Db	1162	-----CTTGCCTCTGATCGAAGATTCCTTATACGGCATCCAGTGAAGTCACTGACTTCGGCTG	1216	
Oy	378	-----	gluGlyIleGlyPheT	383
Db	1217	CGAAGCGACACCTCTCAATATGCCAGATCTTGATAGAAACCATGACAAAGCGCAACTCCG	1276	
Oy	383	hrLLeuTyrGlnGluIleLeuThrGly	-----IleGluProAlaLeuThr	397
Db	1277	ATGTTTACTCCGATGCTACACAGCTCAAGCTCACTAGTACAGCTTCATGCAAAAGCCGGCACTTATA	1368	
Oy	398	-----ProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspI	414	
Db	1337	CGGTGAATATCAAAAGCCGAGGGTTATCTCTGAGGCAACTGTACCATTCATGCATGAAGGCA	1396	
Oy	414	IleGluAlaGluTyrLysSerGluSer	---GlyLeuAsnValGlyTyrSerIleuTyrAsnT	433
Db	1397	AAGAAGCCGTATCATGCATTCATTCGATTTGGGCAACTCGTTCCCTGCTGCTTACCCGAT	1456	
Oy	433	hrGlyGluGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluV	453	
Db	1457	TCACAGCTTCTCTATGACACATCTCAGTAGGCGAAAGGCTCAATTCGAAT	1509	
Oy	453	allIleGluValLysThrSerSerIleAsnIleSerTrpTyrGlyLysGlnHisProG	473	
Db	1510	-----CAAGCACAATATCAACCCAGCAAGTTGGAGTGGAGCTTGCGAAGCGGACAGCT	1563	
Oy	473	IuSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuT	493	
Db	1563	-----	1563	
Oy	493	yrArgArgThrGlyThrGlnGlnIleTrpGluProValArgHisAlaGln	---GlyGlyTyrV	512
Db	1564	-----GCCATGATGATACAGAGCAGATCCGCTGATGCTTAAATGATCATCCCGTCAGTACG	1618	
Oy	512	alaAsnSerIleLeuVal	---AsnThrThrAspProAsnValVal	527
Db	1619	ACGTTTACGCTCAAAATGTGGATGTGATGCACATGGTCTCCACACAGATTCGAAAGAAATTC	1678	
Oy	527	alThrValAspAsnAsnGlnGluLysLeuSerIleValPro	---AsnSerPheValAlaA	546
Db	1679	TCACGTCTCAAT	-----GCCGTATGCCGTACACTCAATTCGCGGTA	1720
Oy	546	spleuAsnSerTyrGlnHisSerThrIleThrValGlnPheAsnSerAspSerProAspG	566	
Db	1721	CCCCGACGGAATATGAAAGAGGCGCAG	---ACGGTATCTTCCAAACCAATCCACACATG	1777
Oy	566	IuIleArgThrProValAlaPheAlaLeuSerThrGlnGlyAlaIleThrAlaAspAspValIleS	586	
Db	1778	CCACCAACCTACGTATGGATATTTCATATGGCGGCACTCCGCTCAACAGTAAAC	1830	
Oy	586	erLeuGlyTyrValMetAlaGluValProGlyGlySerSerAsnTyrProValTyrPS	606	
Db	1831	-----	GAAGACCGCACTGCTTTACA	1853
Oy	606	erLysAspValLeuThrLeuSerGlnGlyAspTyrThrLeuTyrThrArgPheSerIleA	626	
Db	1853	GCAAA	-----GCCGCGCAATGCAATGTCACGCTC	1881
Oy	626	snaAsnGlnLysAspGluTrpLysLysIleGlySer	---ValSerValLysThrProt	644
Db	1882	-----	AAGCGATACAGTCTCCGCTGAAACGGTGAAGCGAAG	1921
Oy	644	hrGluTyrThrHisProLeuPheGluValAlaGlnHisAsnGlnThrSerThrTyrThrLeuA	664	
Db	1922	AAAAATAC	-----	1929
Oy	664	spMetAlaHisAsnArgValLeuProAspPheThrIleuLysAsnLeuGlyLeuProPheA	684	
Db	1930	-----	ATCACGTCTCAAAAGAGCTCCGCTCCT	1956
Oy	684	snGlyGluLeuValValAlaPheArgGlnThrGlnSerSerGlySerLeuThrAlaA	704	


```

Db 1708 -----GCCATGAGTACAGAGCAGAAATCCGCTGTATCCATATAGTCTATCCCGGTCATAGC 1762
QY 512 aIAsnSerIleuVal---AsnThrThraSPProAsnAsnValVal-----V 527
Db 1763 ACGTTACGCTCAAAATGAGTGAAGAGTGTCCACAGATTCAGAAAGAAATATCA 1822
QY 527 aIThrValAspAsnAsnGluGlyLysLeuSerIleValPro---AsnSerPheValAla 546
Db 1823 TCACGTGCAT-----GCCGTATGCTCTAGCTGATTCGTTGCGTGA 1864
QY 546 sPLeuAsnSerTyrgLuhISerThrIleThrValGlnPheAsnSerAspSerProAspG 566
Db 1865 CCCCAGACGGAATAGAGAGGGCCAG---ACGGTATCTTCCAAAACCAATCCACCAATG 1921
QY 566 IuIThrGThProValAlaPheAlaLeuSerThcGlyAlaThrAlaAspAspValIles 586
Db 1922 CCACCACTACCTATGATATTCGATGGCGGCACTCCCGCTCCAGTGAAGAC----- 1974
QY 586 eTLeuGlyTrpValMetAlaGluValProGlyGlySerSerAsnTyrrProValValTrps 606
Db 1975 -----GAAACCCGACTGTGCTTTACA 1996
QY 606 eTlysAspValLeuThrLeuSerGluGlyAspTyrrThrLeuTrpTyrrArgPheSerIleA 626
Db 1997 GCAAA-----GCCGCGCAATAGATGTCACGCTC----- 2025
QY 626 sNAsnGlnLysAspCluTrpLysIleGlySer-----ValSerValLysTrpProt 644
Db 2026 -----AAGCGCATCAGTGTCTCCGCGTGAACGGTGAAGACGAAAG 2065
QY 644 hrGluTyrrThraSPProLeuPheGluValGlyHisAsnGlnThrSerThrTyrrThraLeuA 664
Db 2066 AAAAAATAC----- 2073
QY 664 sPMeAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeuGlyLeuProPheA 684
Db 2074 -----ATCAGTGTCAAGAAAGTCCGGTCCCT----- 2100
QY 684 snGlyGluLeuValValPheArgGlnThrGlnSerSerGlySerLeuTrpAlaA 704
Db 2101 -----GCTCCGCTACCGCAAGGAAACCTCGAATA----- 2136
QY 704 IagGlnGluThrValHisIleLysGlnGlyLuhThrPheValTyrrLysPro----- 720
Db 2137 -----GTAAAGAAAGCGGAGACAGTACTTCAAAAGCTGTGTACGA 2179
QY 721 -----ValValGluGlyProIlePro----- 727
Db 2180 ACAATCCGACTCATGCGTTGGGTTCGAAAGCGGCGCTCTCTGCCACACAGCGAGC 2239
QY 728 -----AspGlySerTyrrArgAlaThrLeuHisAlaPheV 739
Db 2240 AAAACCGGTGTCTACATCAATGAACAGCAAGCATGATCCAGCTGACGTCCACCA 2299
QY 729 aIAsnGlyGlnGlnGlnLeuTyrrLysGlyLysArgAsnTyrrThrValLysIleValA 759
Db 2300 ACGAGGGGCGAAGCAAT-----GTGAAGAAAGCAGAAACATPAC--ATTGAGGTATCC 2350
QY 759 snGlyThrAlaValAlaIleGluSerSerGluGluLeuArgValPheProAsnProA 779
Db 2351 TCGATGACAGTGTGCGACATAGGCGACAGACGGGTATCGTCATCTCGCAAAAG 2410
QY 779 IArgAspTyrrValGluIleSerAlaProCysIleProGlnGluThrSerIleIleLeuP 799
Db 2411 GAACGAAGCAGATCTCTATGAAAGCAACGCTGCTATCAAA-----GGATTCGTTCTCT 2464
QY 799 hAspLeuSerGlyLysIleValMetLysAsnSerLeuAlaGlyHisGlyArgMetA 819
Db 2465 ATGACATCATAGGAGGGGTGACTCAAAACTACCGCATACGCTCCGTCGACCGTAg 2524
QY 819 sPValSerArgLeuProAsnGlyAlaTyrrIleLeuLysVal-----AspGlyTr 835

```

```

Db 2525 ATCTTCCATCTCGCCGGAAGAAATCTACACCATCATATCAAAAGCAAAATCCGCTC 2584
QY 835 yIThrThraSPProIleAsnIle 841
Db 2585 GCACGGAAAGATCATATC 2604

* RESULT 13
ID AAX91740 standard; DNA; 2763 BP.
XX
AC AAX91740;
XX
DT 25-AUG-1999 (first entry)
XX
DE Porphyromonas gingivalis protein PG57 encoding DNA.
XX
KW Porphyromonas gingivalis; PG: periodontal disease; gingivitis;
XX vaccine; antigenic; ds.
XX
OS Porphyromonas gingivalis.
XX
PN W0929870-A1.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98NU-AU01023.
XX
PR 04-AUG-1998; 98AU-0005028.
XX 10-DEC-1997; 97AU-0000839.
XX 31-DEC-1997; 97AU-0001182.
XX 30-JAN-1998; 98AU-0001546.
XX 10-MAR-1998; 98AU-0002264.
XX 09-APR-1998; 98AU-0002911.
XX 23-APR-1998; 98AU-0003128.
XX 05-MAY-1998; 98AU-0003338.
XX 22-MAY-1998; 98AU-0003654.
XX 29-JUL-1998; 98AU-0004917.
XX
PA (CSLC-) CSL LTD.
XX
PI Agius CT, Barr IG, Hocking DM, Margolis MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
DR WPI: 1999-385613/32.
XX P-PSDB: AAY34522.
XX
PT Antigenic Porphyromonas gingivalis peptides for preventing
XX gingivitis
XX
PS Claim 12: Page 235-236; 588pp; English.
XX
CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91889 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
SO Sequence 2763 BP; 767 A; 580 C; 666 G; 750 T; 0 other;

Alignment Scores:
Pred. No.: 0.0015 Length: 2763
Score: 159.50 Matches: 152
Percent Similarity: 31.14% Conservative: 108
Best Local Similarity: 18.20% Mismatches: 231
Query Match: 3.59% Indels: 345
DB: 20 Gaps: 38

US-10-030-330-1 (1-843) x AAX91740 (1-2763)

```

```

OY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyIleuMetTyraSPTrIleAsnmet 240
Db 643 GGTGAGAAATGTGGACGATGACTGTGCAATCTCGAGAGATAT-----CA 650
OY 241 Pro-----GlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyr 257
Db 651 CCGAGAGAAAGGGAATGAGAGAGTTG---CAGCTTGGCGGTACACATCTATGCGAAC 747
OY 258 AlaThrLeuMetAlaGlyAspValSerAlaSerValSerMetSerPheThrGlyIleuGlnGlySer 277
Db 748 GGTACACTACTGGCAAAATATAA-----GATGTCCTCCATCTGAGATGATGGAC 798
OY 278 GlyThrTrpSerValTyrValValGlyAlaLeuArgAsnAspPheArgTyrIleuGlnGlySer 297
Db 799 AGCACTACTCT-----TTGCGAGACAT-----CCC 825
OY 298 LeuGlnLeuHis-----ValArgAlaLeuTyrThrSerGlnGlyTrpHisAspMetIle 315
Db 826 TTGCAAGTGGATGACGCTTACAGCGCTTAC-----858
OY 316 ArgGlyGlnLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnGlnSerIleGly 335
Db 859 -----873
OY 336 HisAlaPheValCysAsp-----GlyTyrAlaSerAspGlyThrPheHisPheAsnTrp 353
Db 874 TCTTGACCGTATGTGGCAGCGCTGATTCACCGAGTGGCATCTTATGAAAATTT 933
OY 354 GlyTrpGlyValSerAsnGlyPheThrGlyLeuThrLeuLeuSerProThrSerLeu 373
Db 934 GAGATGGACCTGTTCCCAATGGTTGG-----CTT 963
OY 374 GlyIleGlyGlyGlyIleGlyPhe-----382
Db 964 GTGATAGACGCTGATGGAGATGATTAAGTGGGACATTTGAATGATACAGACGCT 1023
OY 383 -----ThrlleTyrGlnIleuIleThrGlnIleuIleThrGlnIleu 393
Db 1024 TTTCGGCGCCATTAATGAGCCATGCTGCTGCTGCTGCTTATGTTCCGGGTATAGGC 1083
OY 394 ProAlaLys-----ThrProAlaGlnAlaGlyThrAspAlaLeuPro 407
Db 1084 CCGGTGACTCCGACACATCTGATTAACCCCAAGCTTGAAGGCCAAMCGTGTCAAG 1143
OY 408 Ile-----LeuAlaLeuLysAspIleGlu-----AlaGluTyrLysSerGlySerGlyLeu 424
Db 1144 TACTGGGTAGACACGAGATGCCAATTGGCGAGCAACATTAACGGGTGATGCTTCG 1203
OY 425 AsnValGlyTyrSerIleThrAsnThrGlyGlnGlnSerAsnLeuAspLeuGlyTyr 444
Db 1204 ACAACGGGAGCTGCTGC-----1221
OY 445 ArgLeuAsnLysAlaAspGlyValIle-----GluValLysThrSerSer 460
Db 1222 -----GGAGATTGCTCATATTGTTGAAAGAACATGACAGACGAG 1263
OY 461 IleAsnIleSerTrpTyrGlyTyrGlyGluHisProGlnSerPheSerLeuAlaProAs 480
Db 1264 CCGACCGCGGATGCTAT-----GAAAGAACATCAACTTACTGGAAGGACTAA 1314
OY 480 GlnLeuSerGlnGlyIleAsnThrIleThrLeu-----491
Db 1315 TACATCCGATGCGCGCTTACACTGATATATATTCTTGAAGTTGACGATATC 1374
OY 492 -LeuTyrArgArgThrGlyThrGlnIleTrpGlyProValArgHisAlaGlnGlyTyr 511
Db 1375 ACTGTATTTCGGG-ACCTCCGATCAGAGCCGAACTGTT-----1413
OY 511 ValAsnSerIleLysValAsnThrThrAspProAsnAsnValValIleThrValAspAs 531
Db 1414 -----ACCGAT-----TTGCTGCTGCTTATTGAAA 1442
OY 531 AsnGlnGlyLysLeuSer-----IleValProAsnSerPhe-----543

```

```

Db 1443 CAACAGGAGAGATTAAAGTGAATTTCTTAAGCGCTACGAAACCCGATAGACTGATGA 1502
OY 544 -----ValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGly 558
Db 1503 TAAAGCCCATTCGAGCTTCCCGCTACATATCTGCAAAAGCGCTGCTGTTCA 1562
OY 558 nPheAsnSerAspSerProAspIleArgThrProValAlaPheAlaLeuSerThrGly 578
Db 1563 CATACAAAGCCG-----ACTGTTTGGAGTATCATGATGACACTTA 1604
OY 578 ValAlaThrAlaAspAspValIleSerLeu-----587
Db 1605 TTCTTCAGAGAGATCAGATGAGTGAAGTGAATATGTGTACCTGCCGTTTAAAGCA 1664
OY 587 -----587
Db 1665 TATCAGTCCCAATCGGTTTGGATAGACTATTAAGTATTCATCGACATTAATCTT 1724
OY 588 -----GlyTrpValMetAlaGlnValProGly 596
Db 1725 ATATGAGCGCTTTGAGCGCGAGATATCTGAAAGCGTGTGTTGATGATCTGATGG 1784
OY 596 yGlySerSerAsnTyrProValValTyrSerLysAspValLeuThrLeuSerGlnLys 616
Db 1785 CGACAAATGTAAATGG-----GA 1802
OY 616 PTVThrLeuTrpTyrArgPheSerIleAsnGlnLysAspGluTrpLysIleGly 636
Db 1803 CTATTATCTCTGACTATGATGGA-----CATGACAGAGAACGTATATTCG 1850
OY 636 ySer-----ValSerValLysThrProThrGlyTyr---ThrHisPr 649
Db 1851 ATCCCTTCCTACTTACCGATGATGCTTAACTCCGGATTAATTTGGTTACAC 1910
OY 649 OleuPheGlnValGly-----HisAsnGlnThrSerThrIleTyrTh 662
Db 1911 CAGACTCGAGAGAGCCAGCTTGTCAAGTATGGGTAAAGTGGCAAGATCGTGTATTC 1970
OY 662 rLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeuGlyLeuPr 682
Db 1971 GGCTGAG-----CATATGCTGTGATGCTTCTACTAG-----GGAATGCG 2012
OY 682 oPheAsnGlyGlnLeuValValPheArgGlnThrGlnSerSerGlySerLeuTr 702
Db 2013 TGTGAA---GATTTGCTCTCTGTTGGAAGACATGACCGCTAAGGCTAACGCTGC 2069
OY 702 PalAlaGlnGlnThrValHisIleLysGlnGlyGluThrPhe-----716
Db 2070 ATGTATAGCGAACTATTAATTCCTCGACGAAACAAATATATTTGCTGCGGCATTA 2129
OY 717 -----ValTyrLysProValVa 722
Db 2130 TGATTGACCGATATGTTTCTTCTGCTTGTGATGACATTAACGTTTATCGTTTACGTA 2189
OY 722 IglnGlyProIleProAspLysSerTyrArgAlaThrLeu-----735
Db 2190 GACTGTCCGAGCGCTGTACTGATTTCTGTTGCTGCTTATTAAGAAATACAAAGGTCG 2249
OY 735 -----735
Db 2250 CTTGAATGATTAATCTTAACGCTACAGAACCCGATAAGACTATGATTAACCAATTT 2309
OY 736 -----HisAlaPheValAsnGlyGlnGlnGlnLeuTyrLys-----748
Db 2310 GCAGCTTACCGGCTACAAATCTATGCAAAATGCTGCTCTTGTACATCAAGACCC 2369
OY 748 -----748
Db 2370 GACTGTTTGGATATATGATGATGACTTATTTCAAGAGACGTCAGGTGGAATGCA 2429
OY 749 -----GlyLysAr 751

```

Db 2430 ATATTGTCTACTGCCGTTTATACGACATATCGATCCGATCGTTGGCATAGCT 2489
 QY 751 GASNHYRTHVALVSLILEVALASNGLYRHALAVALIGLUALILEGLUSERSERLUGL 771
 Db 2490 GAACATATCTACTATC-----ACATCCCTGGATTAATATTCATATCTGATACAAAG 2534
 QY 771 ULLEATGVALPHEPROASNPROLAARGSPYRVALIGLUILESERALEPROCYSLLEPR 791
 Db 2535 CTTGAAAATATATCTCAATCCGCAATCGCATGTGTGTAGAGATAGAGGA---TTGAGTCG 2591
 QY 791 OGINGLUTHRSERILEILELEUPHEASPLEUSERGLYLSILEVALMETLYSANSERIE 811
 Db 2592 GAGCAAGTCGACAAATCGATGTATATATCGCGGGAATTTGCATATTTAAGGAGAGAC 2651
 QY 811 USERALEGLYHLSGLYARGMETASPYALSERARGLEUPROASNGLYALATYRILEU 831
 Db 2652 TCATTCAGAGAAAGCAATGATGATTTACGTCATCAATGACGGGCTACTTGTATTA 2711
 831 SYALASPGLY-----TYRTHRILYSILEANILE 841
 2712 ACTAGTCGCTGGAATTAACAACAACCAACCAAGATGAGATTA 2754
 RESULT 14
 AAX91739 standard; DNA; 2766 BP.
 AAX91739;
 AAX91739;
 25-AUG-1999 (first entry)
 Porphyromonas gingivalis protein PG57 encoding DNA.
 Porphyromonas gingivalis; pg; periodontal disease; gingivitis;
 vaccine; antigenic; ds.
 Porphyromonas gingivalis.
 WO9929870-A1.
 17-JUN-1999.
 10-DEC-1998; 98MU-AU01023.
 04-AUG-1998; 98AU-0005028.
 10-DEC-1997; 97AU-0000839.
 31-DEC-1997; 97AU-0001182.
 30-JAN-1998; 98AU-0001546.
 10-MAR-1998; 98AU-0002264.
 09-APR-1998; 98AU-0002911.
 23-APR-1998; 98AU-0003128.
 05-MAY-1998; 98AU-0003338.
 22-MAY-1998; 98AU-0003654.
 29-JUL-1998; 98AU-0004917.
 (CSLC-) CSL LTD.
 PI Agus CF, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 DR WPI; 1999-385613/32.
 DR P-PSDB; AAY34521.
 XX Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 PS Claim 12; Page 234-235; 588bp; English.
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY4318 to
 CC AAY4583. AAX91802 to AAX9189 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can

CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 XX
 SQ Sequence 2766 BP; 768 A; 580 C; 667 G; 751 T; 0 other;
 Alignment Scores:
 Pred. No.: 0.0015 Length: 2766
 Score: 159.50 Matches: 152
 Percent Similarity: 31.14% Conservative: 108
 Best Local Similarity: 18.20% Mismatches: 231
 Query Match: 3.59% Indels: 345
 DB: 20 Gaps: 38
 US-10-030-330-1 (1-843) x AAX91739 (1-2766)
 QY 221 GLYSERLEVALIGLYANSRTPSERGLYTHRPHGLYGLUMETYRASPTRILEASMET 240
 Db 646 GGTGAGAGTGTGGAGATGATGCTTGAATATCCGAGGATAT-----CNA 693
 QY 241 PRO-----GLYSNPPOASPLEUASPNLEUTHRGINSERGLNVALASPALATYR 257
 Db 694 CCGAAGAAAGGGAATGAAGAGTTG--CACTTAGCGGCTACCAACATCTATGCGAAC 750
 QY 258 ALATHREUMETARGSPYALSERALESERVASERSETSERPHERYGLUANGLYSER 277
 Db 751 GGTACACTACTGCGCAAAATTA-----GATCTCCACTACTGCGAGTATGCGAC 801
 QY 278 GLYTHRYTSERVALTYRVALVALIGLYALALEUARGASNPHEARGTYRLYSARGSER 297
 Db 802 AGCACTTACTCT-----TTGCGAGACAAAT-----CCC 828
 QY 298 LEUHLNLEUHS-----VALARGALALEUTYRTHSERGLNGLUTRPHISAPMETILLE 315
 Db 829 TTGCAAGTGGAGTACTGCTGTACAGCGCTTAC-----GATGAACGATGAA 876
 QY 316 ARGGLYGLULLEUALASERGLYARGPROVALTYRVALAGLYSANAENGINSERILEGLY 335
 Db 862 -----GATGAACGATGAA 876
 QY 336 HISALAPHEVALCYSASP-----GLYTYRALASERASPYLYTHRPHHISAPHEANTRP 353
 Db 877 TCTTCACCGTATGTGGACAGCTGATACGATACGCCAGGATCCATTCCTTATGAAAATTTT 926
 QY 354 GLYTRPGLYGLYVALSERASNGLYPHEPTYRLYSLEUTHREUENUSERPROTHRSERLEU 373
 Db 937 GAGAAATGACCTGTCTCCCAATGCTTG-----CTT 966
 QY 374 GLYILEGLYGLYGLULGLYILEGLYPHE-----382
 Db 967 GTGATAGACGCTGATGAGATGATTTAGCTGGGACACATTTGAATCATACGACGCT 1026
 QY 383 -----THRILETYRGLNGLUILEIHRNGLYILEGLU 393
 Db 1027 TTTCCCGGCATATATGAGAGCCATGTGCTCTGCGCTCTTATATGTCGGGATATAGGC 1086
 QY 394 PROALALYS-----THPROALAGLUALAGLYTHRASPALALEUPRO 407
 Db 1087 CCGGTGACTCCGACACATATCTGATTAATCCCAAGGTTCGAAGCAACGTCGCAAG 1146
 QY 408 ILE--LEUALALEUYSAPILEGLU-----ALAGLUTRYRYSERGLUSERGLYLEU 424
 Db 1147 TACTGGTATAGCGACGAGATGCCAATTTGGGACGCGGACATTAACGGGTGATGCTTCG 1206
 QY 425 ASNVALGLYTYRSEIRLETYRASNTHRGYGLUGLUGLINSERASNLEUASPLEUGLYTYR 444
 Db 1207 ACAAGGGGACCTCGTGC-----1224
 QY 445 ARGLEUASNLYSALASPGLYGLUVALILE-----GLUVALYSTRHSER 460
 Db 1225 -----GGAGATTTGCTCATATTTGTCGAGAACCAATGACGAGAG 1266
 QY 461 ILEASNLISERTPYRGLYTYRGLYGLU-HISPROGLUSERPHESEIRLEUALAPROAS 480

```

Db      1267 CCGACCGCGCGATGTTAT-----GAAAGAACCACTACCTAGAGGACTAAA 11
Oy      480 nglneuserelneilylleasntrhlethleu----- 1317
Db      1318 TACATCGCATGCGCGCATACACTGATACCGATATATATTCTTGAAGTTGACGATATC 491
Oy      492 -leutyfargtrgthrglythruglntprgluprovalarghslagllyglyty 1377
Db      1378 ACTGTATTGCGG-ACCTCGCATCGAGCGCCGAACTCGTT----- 1416
Oy      511 rvalsanserilleysvalasnthrthrasproasnvalvalthvalaspas 531
Db      1417 -----ACCGAT-----TTCGTGTCTCGCTTATTTGAAA 1445
Oy      531 nasngluglylylseuser-----llevalproanserphe----- 543
Db      1446 CAACAAGGACGATTAAAGTGAATTATCTCAAGCGTACGACACCGATAGACTGATGA 1505
Oy      544 -----ValAlaAspLeuAsnSerTyrgluHisSerThrIleThrValG1 558
Db      1506 TAAAGACCATTCGACGCTTCCGCGCTACAAATATCTACCAACGCGCTCCTGTGTTCA 1565
Oy      558 nheasnseraspsersproasgluileargthrprovalalaphaalalseuserthrg1 578
Db      1566 CATACAAAGACCG-----ACTGTTTGGAGTATATCGATGACTTA 1607
Oy      578 yalathralaaspaspyallieserleu----- 587
Db      1608 TTCTTACGACGACCATGAGGTGAGATGATATGTGTCTACGCGGTTATACGACAA 1667
Oy      587 ----- 587
Db      1668 TATGAGTCCCAATCGGTTTCGATAGCTGATTATGATTCCTCAATCGGACATTATCTT 1727
Oy      588 -----GlyThrValMetAlaGluValProG1 596
Db      1728 ATATGAGGCTTTGAGCGCGGAAGTATCTCTGAAGCGTGGTTGATTGATGCGATG 1767
Oy      596 yglseraseraspytyprovalvaltyrserlyaspyalleuthrleusergluglyas 616
Db      1788 CGACAACTTTAAATGG-----CATGACAGTGAAGAGTATTCG 1805
Oy      616 ptythrleutpytyrargpheSerlleasnagllyaspyallutpplyslylleg1 636
Db      1806 CTATTATCTCTGACTATGATGGA-----CATGACAGTGAAGAGTATTCG 1853
Oy      636 yser-----ValSerVallysthrProthrglutyrt---ThrHispr 649
Db      1854 ATCCCTTCGTACTTACCGATGATGCGGTTTAACTCCGATACATATTGGTTACACC 1913
Oy      649 oleuphegluvalgly-----Hisasnglnthrserthtyrth 662
Db      1914 CAGACTCGAAGGACCAAGCTTGTCAAGTATGGGTAAAGTGCAGCAAGATGCTTTATTC 1973
Oy      662 rleuaspmetalahisasnrgvalleuproasphethrleuysasnleuglyleupr 682
Db      1974 GCGTGAG-----CATTAGCTGTGATGTTTCTACTACG-----GCACTGC 2015
Oy      682 ophasnrglygluleuvalvalpheaarglnthrgrlnserSerlyserleutyr 702
Db      2016 TGTGAA-----GATTTGTCTCTCTTGTTCGAAGACATGACCGCTAAGCGTACG 2072
Oy      702 palalaaglnluthrvalhisilleysglnclyluthrph----- 716
Db      2073 ATGCTATGAGCAACTATTACTTCCTGCGAGAACAAATATATTCCTGCGCGCATTA 2132
Oy      717 -----ValIlyrlyspProvalva 722
Db      2133 TGATTGCAACGATATGTTTCTTGTCTGTGATGACATTAACGTTATGCTTACTGA 2192
Oy      722 lsluglyproleuproaspylysetyrfargalathrleu----- 735

```

```

Db      2193 GACTGTCCGAGCGCTGTACTGATTTGTTGTCGCTTATTAGAAATACAGGTCG 2252
Oy      735 ----- 735
Db      2253 CCTGAATGGAATATATCTTACCGCTACAGACCCGATAGACTGATGATAAAAACATT 2312
Oy      736 -----HisAlaPheValasnlyglnglntglntleuTyrlleuys----- 748
Db      2313 GCAGCTTACCGGCTACACATCTATGCAAAATGCTGCTGTCTGTTCAATACAAAGCC 2372
Oy      748 ----- 748
Db      2373 GACTGTTTGGATATATGATGATGACTTATTCTTACGAGACGCTGAGTGAATGGA 2432
Oy      749 -----Glyysar 751
Db      2433 ATATTGTGCTACTGCCGTTTATACGACAAATATGATCCCAATCGGTTGCAATAGCT 2492
Oy      751 gAsnTyThrVallyslleValasnlythrAlaValGluAlaIleGluSerSerluG1 771
Db      2493 GAACTATCTATC-----ACATCTTGATATATATCAATCTGATCAAG 2537
Oy      771 uileargvalpheproasnproalaargaspytyrfvalgluileserAlaProCysIlepr 791
Db      2538 CTGAAATATATCTTAATCCGACATCGATATGTGTAGAGATAGAGGA---TTGAGTCG 2594
Oy      791 oglnluthrSerlleleuPheaspleuserglylyslleValMetlyasnserle 811
Db      2595 GAGCAAGTCGACAAATGAGTTGTATATGCGTGGGAATTTCCATATTAAGGAAGAC 2654
Oy      811 useralaglyhisglarymetaspyalserargleuProasnlylatyrylleuTy 831
Db      2655 TCATTGAGAAAGAAAGCAATATGATTTTACGCTCATGACGAGTCTACTGATTAA 2714
Oy      831 sValaspyly-----TyThrThrylleasnle 841
Db      2715 AGTAGTCGTTGCAATTAACAAACAAACGAAAGGTAGATA 2757
Oy      2757 -----
RESULT 15
AAx91738 standard; DNA; 2775 BP.
ID AAx91738
XX AAx91738;
AC AAx91738;
DT 25-AUG-1999 (first entry)
DE Porphyromonas gingivalis protein PG57 encoding DNA.
KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic; ds.
OS Porphyromonas gingivalis.
XX MO9929870-A1.
PN 17-JUN-1999.
PD 17-JUN-1999.
PE 10-DEC-1998; 98WC-AU01023.
PF 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX (CSLC-) CSL LTD.
XX Aglus CT, Barr IG, Hocking DM, Margetis MB, Patterson MA.

```



```

Db 1983 GGCTGAG-----CATTTAGTGTGATGTTCTACTACG-----GGAAGTGC 2024
Oy 682 ophesanglyluLeuValValPheArgGlnThrGlnSerSerGlySerLeuTr 702
Db 2025 TGTGAA---GATTTTGTCTCTTTGTTCGAGAGACATGACCGCTAAGGCTACGGTGC 2081
Oy 702 palaaaglngluThrValHisIleLysGlnGlyLeuThrPhe----- 716
Db 2082 ATGTATGAGCAACTATTACATTCGCTGCAGAACAAATATATTGCTGGCGGCAATTA 2141
Oy 717 -----ValTyrLysProValVa 722
Db 2142 TGATTGACCGCATATGTTTCTGCTCTTGATGACATTAAGCGTTTATCGTTTACTGCA 2201
Oy 722 LgluLyrProIleProAspGlySerTyrArgAlaThrLeu----- 735
Db 2202 GACTGTCCCGAGCCTGTTACTGATTTGCTGTCTGCTTATTAGAAATACAGAGGTGC 2261
Oy 735 ----- 735
Db 2263 CCTGAATGATATATCTAACGGCTACGACCCGATAGACTGATGATAAAGCAATT 2321
Oy 736 -----HisAlaPheValAsnGlyGlnGlnLeuTyrLeuLys----- 748
Db 2322 GCAGCTTACCGGCTACACATCTATGCAATGCTGCTGCTTGCATACACAGACC 2381
Oy 748 ----- 748
Db 2382 GACTGTTTGGATATATCGATGAGACTTATCTTCACAGACGCTCAGTGGAAATGA 2441
Oy 749 -----GlyLysAr 751
Db 2442 ATATTGTGCTACTGCCGTTTATACGACAAATATCCAGTCCCAATCGGTTTGGCAATAGCT 2501
Oy 751 gAsnTyrThrValLysIleValAsnGlyThrAlaValGluAlaIleGluSerGluGlu 771
Db 2502 GAACATATCTATC-----ACATCTCTGGTGAATATTCATTCATTCGATACAAAG 2546
Oy 771 uileArgValPheProAsnProAlaArgAspTyrValGluIleSerAlaProCysIlePr 791
Db 2547 CTGAAATATATATCTTAATCCGCGATCGTATGTTGTAAGGATAGAGGCA---TTGAGTGC 2603
Oy 791 oGlnGluThrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLysAsnSerIe 811
Db 2604 GAGCAAGTCGACAAATCGACTGTATATATCCGCTGGGAATTTCCATATTAAGGAGAGAC 2663
Oy 811 uSerAlaGlyHisGlyArgMetAspValSerArgLeuProAsnGlyAlaTyrIleLeuLys 831
Db 2664 TCATTTCAGAGAAACGGAATATGATGTTTCACGCTCAATGACGGAGTCTACTTGATTA 2723
Oy 831 sValAspGly-----TyrThrTyrLysIleAsnIle 841
Db 2724 ACTAGTCGCTGGAATTAACAAACAAACGAAAGATAGAGATA 2766

```

```

Pf 10-DEC-1998; 98WO-AU01023.
xx
Pr 04-AUG-1998; 98AU-0005028.
Pr 10-DEC-1997; 97AU-0000839.
Pr 31-DEC-1997; 97AU-0001182.
Pr 30-JAN-1998; 98AU-0001546.
Pr 10-MAR-1998; 98AU-0002264.
Pr 09-APR-1998; 98AU-0002911.
Pr 23-APR-1998; 98AU-0003128.
Pr 05-MAY-1998; 98AU-0003338.
Pr 22-MAY-1998; 98AU-0003654.
Pr 29-JUL-1998; 98AU-0004917.
xx
Pa (CSLC-) CSL LTD.
xx
Pi Agus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
Pi Ross BC, Rothel LJ, Webb EA;
xx
Dr WPI; 1999-385613/32.
xx
Dr P-PSDB; AAY34392.
xx
Pt Antigenic Porphyromonas gingivalis peptides for preventing
Pt gingivitis
xx
Ps Claim 12; Page 146-147; 588pp; English.
xx
Cc AAY91536 to AAY91801 encode two hundred and sixty six antigenic
Cc Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
Cc AAY34583. AAY91802 to AAY91989 represent PCR primers used in the
Cc isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
Cc activity with a vaccine mechanism of action. The Pg polypeptides can be
Cc used as vaccines especially against Porphyromonas gingivalis. Probes can
Cc be used to detect Porphyromonas gingivalis in standard hybridisation
Cc assays. Porphyromonas gingivalis is involved in periodontal disease
Cc especially gingivitis.
xx
Sg Sequence 2814 BP; 791 A; 586 C; 674 G; 763 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 0.00154 Length: 2814
Score: 159.50 Matches: 152
Percent Similarity: 31.14% Conservative: 108
Best Local Similarity: 10.20% Mismatches: 231
Query Match: 3.59% Indels: 345
DB: Gaps: 38

```

US-10-030-330-1 (1-843) x AAY91610 (1-2814)

```

Oy 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsnMet 240
Db 694 GGTCAAGAAATGTGGGACGATTCGTAACCTATCCGAGGATTA-----CAA 741
Oy 241 Pro-----GlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyr 257
Db 742 CCGAAGGAAAGGAAATGAAAGAGTTG---CAGCTTAGCGGCTACAAACATTCATCGAAG 798
Oy 258 AlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTyrGlnAsnGlySer 277
Db 799 GGTACACTACTGCGCAATAATAA-----GATGCTCATACTGAGATATGTGAGC 849
Oy 278 GlyThrTyrSerValTyrValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSer 297
Db 850 ACGCACTTACTCT-----TTGCGAGACAAAT-----CCC 876
Oy 298 LeuGlnLeuHis-----ValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIle 315
Db 877 TTGCAAGTGGAGTACTGCTTACACCGCTTAC----- 909
Oy 316 ArgGlyGluLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGly 335
Db 910 -----GATGAAGCATAGAA 924
Oy 336 HisAlaPheValCysAsp-----GlyTyrAlaSerAspGlyThrPheHisPheAsnTrp 353

```



```

Db      925 TCTTGACCGGTATGGCAGCGCTGCATTCAGCCAGCATCCTTATGAAATTTT      984
Qy      354 GYTPRGLYGLYVALSERASNGLYPHERYLYSLEUTHLEUSERPROTHRSERLEU 373
Db      965 GAGATGAGCGCTGTTCCCAATGGTTGG-----CTT      1014
Qy      374 GYLIEGLYGLYGLYGLYILEGLYPHE-----382
Db      1015 GTGATAGACGCGATGAGATGAGATTTAGCTGGGCACTATTGATCAACAGCGCT 1074
Qy      383 -----ThrIleTYRGLNGLIleIleThrGLYILEGLI 393
Db      1075 TTTCCCGGCATATGAGAGCCATTGCTCTGCGGCTTCTATGCTCCGGGTATAGGC 1134
Qy      394 PROALALYS-----ThrProALAGLALAGLYThrasPALALeupro 407
Db      1135 CCGGTAGCTCCGACACTATCTGATTTACCCCAAGGTTGAAGGAGCCAAAGCTGTCAAG 1194
Qy      408 Ile---LeuAlaLeuLYSAspIleGLU-----AlaGLUTYRlySerserGLYleu 424
Db      1195 TACTGGGTAGCAGCAGCATCCCAATTGGGACGGGAAACATTACCGGTGATGCTTGC 1254
Qy      425 AsnValGLYTYRSerIleTYRnAnThrGLYGLNGLNGLNGLNGLNGLNGLNGLNGLN 444
Db      1255 ACAACGGGACGTGTCTC-----1272
Qy      445 ArgLeuAnLYSAlaAspGLYGLYValIle-----GLYVallySthrsSer 460
Db      1273 -----GGAGATTTCGCAATTTGTCGAAGAAACCATGACGCAAG 1314
Qy      461 IleAnIleSerTYRtyrGLYtyrGLYGLNGLNGLNGLNGLNGLNGLNGLNGLNGLN 480
Db      1315 CCGACGGGCGCATGTAT-----GAAAGACCATCACTTACCTGAAGGACTAA 1365
Qy      480 nGLNLeuSerGLNGLYIleAnThrIleThrLeu-----491
Db      1366 TACATCGCATGGCGCATTTACACTGATCCGATATATTTCTGAAAGTTGACGATATC 1425
Qy      492 -LeuTYRArgArGthrGLYthrGLNGLNGLNGLNGLNGLNGLNGLNGLNGLNGLN 511
Db      1426 ACTGTTTCGGG-ACCTCTGCATCAAGCGCCGACCTGTT-----1464
Qy      511 rValAnSerIleLYSValAnThrThrasProAnAsnValValIleThrValAspAs 531
Db      1465 -----ACCGAT-----TTCGTTGTCCTTATGAAA 1493
Qy      531 nasnGLNGLYLYSLeuSer---IleValProAnSerPhe-----543
Db      1494 CAACAAAGGAGCATTTAAAGTGAATTATCCCTAAAGGCTCAAGAACCGATTAAGCTGATGA 1553
Qy      544 -----ValAlaAspLeuAnSerTYRGLNGLNGLNGLNGLNGLNGLNGLNGLNGLN 558
Db      1554 TAAAGACCCATTGACGCTTGGCGGTACAAATCTGACAAAGCGCTCCCTCTGTTCA 1613
Qy      558 nPheAnSerAspSerProAspGLUleArGthrProValAlaPheAlaLeuSerThrGL 578
Db      1614 CATCAAAACCGG-----ACTGTTTGGAGTATATCGATGAGACTTA 1655
Qy      578 yAlaThrAlaAspAspValIleSerLeu-----587
Db      1656 TTCCTCAGCAGACGATCAGTGAAGTGAATATGTGTCACGCGTTTATAAGCAAA 1715
Qy      587 -----587
Db      1716 TATCAGTCCCAATCGTTTGGATAGCTGATTATGATTCATCGACGACATTAATCTT 1775
Qy      588 -----GlyTrPValMetAlaGLYValProGL 596
Db      1776 ATATGAGGCTTTGAGCGCGAGATATCTGAAAGCTGTGTTGATGATGACGATG 1835
Qy      596 yGLYSerSerAsnTYRProValValITRPSerLYSAspValLeuThrLeuSerGLNGLN 616

```

```

Db      1836 CGACAATGTTAATTCG-----GA 1853
Qy      616 pTYRThrLeuTrpTYRArgPheSerIleAsnAnGLNLYSAspLUTrpLYSlyIleGL 636
Db      1854 CTATTATCTCTGGACTATGATGA-----CATGACAGTGAGAGATGATATGC 1901
Qy      636 ySer-----ValSerValLYSthrProThrGLUTYR---ThrHSPR 649
Db      1902 ATCCCTTCGTACTTACCGATGATGGCGTTTAACTCCGGATTAATTTGGTTACCC 1961
Qy      649 oLeuPheGLYValGLY-----HisAnGLNthrsThrTYRth 662
Db      1962 CAGACTTCAGAGAGCCAGCTGTCAGATTTGGTGAAGTGGCCAGATGCTGTTATTC 2021
Qy      662 rLeuAspMetAlaHisAnArgValLeuProAspPheThrLeuLYSAnLeuLYLeupr 682
Db      2022 GGCTGAG-----CAATTAGCTGTGATGGTTTCTACTACG-----GAACTGC 2063
Qy      682 oPheAnGLYGLNLeuValValAlaPheArgGLNthrsSerSerserGLYSerLeutr 702
Db      2064 TGTGAA-----GATTTGTCTCTTGTTCGAAGACAAATACCGCTTAAGCTAACGTCG 2120
Qy      702 pAlaAlaGLNGLNthrValHisIleLYSGLNGLYGLNthrPhe-----716
Db      2121 ATGCTATGAGCGCACTATATACATTCGCTGCAGAACAAATATTTGCTGCGGCATTA 2180
Qy      717 -----ValTYRlySProValIle 722
Db      2181 TGATTGCACCGATATGTTTCTTCTGCTCTTGATGACATTAACGTTTATGCTTACTGA 2240
Qy      722 lGLNGLYProIleProAspLYSeryrArgAlaThrLeu-----735
Db      2241 GACTGTCCGAGCCGTACTGATTTCTGCTGCTGCTTATGAGATTAACAAGGCTCG 2300
Qy      735 -----735
Db      2301 CCTGAATGGAATTAATCTTAACGCTAAGCAACCCGATTAAGCTGATTAATAACCAT 2360
Qy      736 -----HisAlaPheValAnGLYGLNGLNGLNGLNGLNGLNGLNGLNGLNGLN 748
Db      2361 GCAGCTTACCGGCTACACATTCATGCAATTAAGGCTGCTCTTGTTCATCAACAGACC 2420
Qy      748 -----748
Db      2421 GACTGTTTGGATATATGATGATGACTTATTTCTCAAGAGAGCTGAGTGAATGCA 2480
Qy      749 -----GlyLYSAr 751
Db      2481 ATATTGTGTCACGCGCTTATTAAGCAAAATATCGAGTCCCATCGTTTGCATTAAGCT 2540
Qy      751 gAsnTYRThrValLYSleValAnGLYThrAlaValAlaGLNGLNGLNGLNGLNGLNGLN 771
Db      2541 GAACCTACTATC-----ACATCTTGGAAATATTAATCAATCATATCAACAG 2585
Qy      771 uIleArgValPheProAnProAlaArgAspTYRValGLNGLNGLNGLNGLNGLNGLNGLN 791
Db      2586 CTTGAAATATATATCCATATCCGCGCATCGATGCGTGAAGATGAGAGAGGA---TTGACTCG 2642
Qy      791 oGLNGLNthrsSerIleIleuPheAspLeuSerGLYLYSleValMetLYSAnSerIle 811
Db      2643 GAGCAAGTCGACAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2702
Qy      811 uSerAlaGLYHisGLYArgMetAspValSerArgLeuProAnGLYAlaLeuLYLeuLY 831
Db      2703 TCATTTCAGAAAGCAAGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 2762
Qy      831 sValAspGLY-----TYRThrTYRlySleAsnIle 841
Db      2763 AGTAGTCGTTGGAATTAATAACAACACCAAGGATGAGATA 2805

```

RESULT 17
 AAH81312
 ID AAH81312 standard; DNA; 3978 BP.

XX AAH81312;
AC
XX
DT 21-SEP-2001 (first entry)
XX
DE Escherichia coli protein encoding nucleotide sequence SEQ ID NO:111.
XX
KW Escherichia coli; identification; proliferation; microorganism;
KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW bacterial growth inhibition; ds.
XX
OS Escherichia coli.
XX
PN WO200148209-A2.
XX
PD 05-JUL-2001.
XX
PF 19-DEC-2000; 2000WO-US34419.
XX
23-DEC-1999; 99US-0173005.
XX (ELIT-) ELITRA PHARM INC.
XX
PI Forsyth RA, Ohlsen KL, Zyskind JW;
XX
DR WPI; 2001-457376/49.
XX
P-PSDB: AAG98256.
XX
PT Novel nucleic acids encoding proteins required for Escherichia coli
XX proliferation, useful for screening for antimicrobial agents -
XX
PS Claim 9; Page 148-154; 596pp; English.
XX
XX The present invention describes a purified or isolated nucleic acid
CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC microorganism is capable of inhibiting proliferation of a microorganism.
CC (I) have antibacterial and antibiotic activities, and can be used in
CC gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC of the microorganism, and the manufactured antibiotic is useful for
CC reducing the activity or level of a gene product required for
CC proliferation of a microorganism in a subject, specifically humans. The
CC nucleic acids that inhibit bacterial growth or proliferation can be used
CC as antisense therapeutics for killing bacteria. In addition to
CC therapeutic applications, the nucleic acid sequences complementary to
CC sequences required for proliferation can be used as diagnostic tools.
CC For example, nucleic acid probes complementary to proliferation-regulated
CC sequences that are specific for particular species of microorganisms can
CC be used as probes to identify particular microorganism species in
CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC represent oligonucleotides, which are used in the exemplification of the
CC present invention.
XX
XX
SO Sequence 3978 BP; 1071 A; 819 C; 1143 G; 945 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0.00715 Length: 3978
Score: 154.00 Matches: 159
Percent Similarity: 33.97% Conservative: 89
Best Local Similarity: 21.78% Mismatches: 252
Query Match: 3.47% Indels: 230
DB: 22 Gaps: 40
US-10-030-330-1 (1-843) x AAH81312 (1-3978)
OY 161 ASnglYhIsPhaAlaAsnAspPromeTatgTTPasngInGlyTYrProTTPasnsnlyS 180
DB 883 AATGATATTCGTAATGCGACAGTCATATTTCCATATATGACTGATATATACAA 942
OY 181 GIuProlLeu-----ProAsnglYhAsnHIsAlaTYrThGlycYsValAlaThr 197
DB 943 GAATATTCTTGTGGCGCTTCAAGCAGGCTCCACGGGTGTGTCAACGTGACCGATAA 1002

OY 198 AlaAlaAlaGlnIleMetArgTYrHisSerTTProLeuGlnGlyGIuGlySerPheAsp 217
DB 1002 ----- 1002
OY 218 TYrHisAlaGlySerLeuValGIuAsnTPSer-----GlyThPheGlyGIu----- 233
DB 1003 -----GGCATTCGAATTTCTTCGTAACG---GGCCACATTTTC 1038
OY 234 MetTYrAspTTPiLeAsnMetProGlyAsnProAspLeuAsnLeuThGlnSerGln 253
DB 1039 CCGTATATCTATATGATGCTGATCTGCGCGGTGAATCTATCTCGATGAAGCAAA 1098
OY 254 ValAspAlaTYrAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTYr 273
DB 1099 GTAGAT-----TCGGAAATTATCATCTCGGGGATGAAA 1131
OY 274 GIuAsnGlySerGlyThTYrSerVal-----TYrValAlaGIuAlaLeuArg 289
DB 1132 GAACAGGCGACAGCAGCAATCTATGATGAAGAACCTCCATATCATCTATCTCGGA 1191
OY 290 AsnAsnPheArgTYrLYsArgSerLeuGlnLeuHIsVal-----ArgAlaLeuTYrThr 307
DB 1192 ACTAATCTTGTTATGACGCCACAGCGCAAAATGAATATCATATCAGGGCTGTGTTC 1251
OY 308 SerGlnGluTTPHisAspMetIleArgGlyGIuLeuAlaSerGly----- 322
DB 1252 ACCAACGAGGAAGTTATCATCTCGGTATGAGAAACCGCGCGGAATGTCAGCATACC 1311
OY 323 -----ArgProValTYrTYrAlaGIuAsnAsnInsSerIle 334
DB 1312 ACGGGGGAGATGTGGAGGTCAATAGAAATGCTTATACCACTGCTGTGGCGGCG 1371
OY 335 GIuHIsAlaPheValCysAspGly-----TYrAlaSerAspGIuThrPheHisPhe 351
DB 1372 GGAACCTCAATATATAGCATGCGGTAGTTCGATTCGCAAAATATTTCTTT----- 1425
OY 352 AsnTPGlyTYrPGLyGIuValSerAsnGlyPheTYrLYsLeuThrLeuLeuSerProThr 371
DB 1426 -----TTGGCGCATTAACAGCGGTATCGGCACACTGATGATGATGACGACA 1476
OY 372 Ser-----LeuGlyIleGlyGIuGlyIle----- 380
DB 1477 TCATGCTTCGAACTGTGGGTATCAATGTGGTAAATTTGTAGCGATTCCTAATATGTC 1536
OY 381 -----GlyPheThrIleTYrGlnGlu-----IleIleThrGlyIleGIuProAla 395
DB 1537 ACTAATGTCGCCACCTTATATCAACGGGCTATGATTTATCGAGAAATGCTCCGGT 1596
OY 396 LYsThrProAlaGlnAlaGlyThrAspAlaLeuProIleLeuAlaLeuAspIleGIu 415
DB 1597 AAGGGATAGTAAATATTTCAACGGGACACTCTCGAATTTTAAAGCATCATCTACTAAC 1656
OY 416 AlaGIuTYrLYsSerGIuSerGIuLeuAsnValGIuTYrSerIleTYrAsnThrGIuGIu 435
DB 1657 GCACAAATTC-----CTACAGTCTGGT-----GTATTCAGCACCGGGTGAA 1695
OY 436 GIuGlnSerAsnLeuAspLeuGIuTYrArgLeuAsnLYsAlaAspGIuValIleGIu 455
DB 1696 -----CTGAATATATACACCGGAGGTATAGTTAA 1725
OY 456 ValLYsThrSerSerIle-----AsnIleSerTPyrgly 467
DB 1726 GCGCGTATACACAGTACTCTCATATGACAAAGTAAAGCGAGCTGAGGTGATGGG 1785
OY 468 TYrGIuGlnHIsProGluSerPheSerLeuAlaProAsnGlnLeuSerGIuIleAsn 487
DB 1786 CAGAACTCTCTCTTGAATCAATATATGATAGTAGAGG-----ACATCTTACCGGCT 1839
OY 488 ThrIleThrLeuLeuTYrArgArgThr-----GlyThrGIuGlnIleTP----- 501
DB 1840 ACCTTAACTTACGAAATACGCTAGCTGAATGTGCAAGGTGAGAACTTTACTTAGGT 1899

```

OY 502 -----GIUProval-----ArghSala 507
DB 1900 GTTTTGGCTGCTGAGCAAGCTAAACATTGGTGTCTGCTACGGTACGGCGGCGCA 1959
OY 508 GlnGlyGlyTyrVal---AsnSerIleLysVal----- 517
DB 1960 GATCGCGGATTATATACCAATGCGACGAAAGTGAGTTTGTCTTGGCGAAGCGCTTTT 2019
OY 518 -----AsnThrThrAspProasn-----AsnValValAlaThrVal 529
DB 2020 GTCTTTAATCAATACCAATACAGTATGCGGCTACAGGTCGATGCTGATTACAGGT 2079
OY 530 AspsnaenGlnGlyLysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSer 549
DB 2080 GACGATAAGACGGAAGATGATC----- 2103
OY 550 TyrGlnHisSerThrIleThrValGlnPheAsnSerAspSerProAspGlnIleArgThr 569
DB 2104 -----CATGATGACGAGCCATACGGTGTTCATGAGGAACTTATAGCGTAAACG 2157
OY 570 ProValAlaPheAlaLeuSerThrGlyAla-----ThrAlaAspAspValIleSerLeu 587
DB 2158 CTGGTCAATGACGGCTCTGCTCAATGCGGCTCATGCGGAGATGGGTTACGGGCAATG 2217
OY 588 GlyTrpValMetAlaGluVal-----ProGlyGlySerSerAsnTyrProval 603
DB 2218 GGG-----TCGAGTGAAGTAAACCATGTCAAACCCCGT----- 2250
OY 604 ValTrpSerLysAspValLeu-----ThrLeuSerGlnGlyAspTyrThrLeuTyrPyr 621
DB 2251 -----ACGTCGACATTCGTCATCAACGACGAGTACGAGATTCACGCTGACCAAT 2304
OY 622 ArgPheSer-----IleAsnAsnGlnLysAspGluTrpLysIleGlySer 637
DB 2305 GCGCTCAAGCGCATGCTGATGCGAGCTGATCCTCCGACAAAGATGTTTGGC 2364
OY 638 ValSerValLysThrProThrGluTyrThrHisProLeuPheGluValGlyHisAsnGln 657
DB 2365 TTTACCCATGCAACAGGACTGAAATTCGCC-----GGTTTGGCCCACTGAA 2412
OY 658 ThrSerThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLys 677
DB 2413 GACAGTACCTTCACCTGGAACGCGACAAACCGCTGCGCTTACACGCGATGTTGCAG 2472
OY 678 AsnLeuGlyLeuProPheAsnGlyLysLeuValValPheArgGlnThrGlnSerSer 697
DB 2473 TCT-----GACAGTGAATAATACACATCGGTAAAGTTGGAGACCAATCC 2517
OY 698 SerGlySerLeuTrpAlaAlaGlnGlnThrValHisIleLysGlnGlyGluThrPheVal 717
DB 2518 ATTGGTGCACG-----GCCATGCAATGGAGGTACCATCATTT 2553
OY 718 Tyr-----LysProValValGlnGlyProIleProAspLysSerTyr 731
DB 2554 TTCGATACGATATTCCTGCTGCGACGCTCGCGGAGGATATATC----- 2598
OY 732 ArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnGlnLeuTyrLeuLysGlyLysArg 751
DB 2599 -----ACGTCGATACCTGCTGCTGCGCGCGGTGACTACCTGGAAGAGGC---CGT 2649
OY 752 AsnTyrThrValLysIleValAsnGlyThrAlaValGluAlaIleGluSerSerGlnGlu 771
DB 2650 AACATATCAG-----GTAAACGCGACGGCGGCGATCTTATGAC----- 2688
OY 772 IleArgValPheProAsnProAlaArgAsp 781
DB 2689 -----GTGCCCTAAACCGTGAATGAT 2709

```

RESULT 18

AAH75059

AAH75059 standard; DNA; 4059 BP.

AC AAH75059;

```

XX 29-OCT-2001 (first entry)
DT Nucleotide sequence of xyloglucanase enzyme.
XX xyloglucanase; family 44; glycosyl hydrolase; detergent;
DE cellulosic fiber; textile scouring; ss.
XX Paenibacillus polymyxa.
OS
XX Key Location/Qualifiers
XX CDS 1..4059
XX FT /tag= a
XX FT /product= "xyloglucanase"
XX WO200162903-A1.
XX 30-AUG-2001.
XX 21-FEB-2001; 2001WO-DK00116.
XX 24-FEB-2000; 2000DK-0000291.
XX (NOVO ) NOVOZYMES AS.
XX PA
XX PI Schorr K, Jorgensen PL, Schuelein M;
XX DR WPI; 2001-522819/57.
XX DR P-PSDB; MAG63962.
XX PT New xyloglucanase enzyme belonging to glycosyl hydrolases family,
XX useful for detergent compositions, and textile or cellulose fiber
XX processing industries
XX PS
XX Claim 14; Page 80-81; 97pp; English.
XX
XX The present sequence encodes a xyloglucanase of the invention. The
XX specification describes a xyloglucanase enzyme belonging to family 44
XX of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
XX of at least 30% at pH 5-8. The enzyme exhibits high performance in
XX detergent compositions and prevents binding of certain soils to the
XX xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
XX room temperature and has a half life of more than 50 days when incubated
XX in a full formulated liquid detergent at 30 plus degrees Celsius. The
XX enzyme is used in detergent compositions, textile industry for improving
XX the properties of cellulosic fibers, yarn, and woven or non-woven
XX fabrics, preferably in textile scouring process, and in cellulose fiber
XX processing industry for ratting of fibers e.g. hemp, jute, flax, and
XX linen.
XX
XX Sequence 4059 BP; 1187 A; 893 C; 1046 G; 933 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 0.0214 Length: 4059
XX Score: 148.50 Matches: 176
XX Percent Similarity: 32.10% Conservative: 127
XX Best Local Similarity: 18.64% Mismatches: 334
XX Query Match: 3.35% Indels: 307
XX DB: 22 Gaps: 47
XX
XX US-10-030-330-1 (1-843) x AAH75059 (1-4059)
XX
XX OY 112 AspMetAspSerMetProAspAsnLeuArgMetTrpLeuGlnIleTyrAspGlnIle 131
DB 844 GATTGGGACTCTGTAAGGCAATATACCTGTTCTGAGACTATTATCCTGATCAAAAG 903
OY 132 GlyLeu----- 133
DB 904 CGCCTTAGCTCCCAAGTCAGCAAGCAAGATTCGATGATTCAGCTACAGTACGTTAT 963
OY 134 -----IleLeuSerGlyLysAlaGlnLeuAsnGlnIleLeuArgThrGlnGlyVal 151
DB 964 CCCGAAGCATGGCGGAGGAGCATACGAATTACGAATGAGGTAGCGAATACGAAGCAAG 1023

```

```

OY 152 ProAlaGluValIHisAlaLeuMetCaspAsnGlyHisPheAlaAsnAspProMetArgTyr 171
DB ||| ||| |||
DB 1024 AAAGCCAGATGACAGCA-----CCTGCACACTTGTGG 1056
OY 172 AsnGlnGlyTyr-----ProTyrAsnAsnLys-----GluProLeu 183
DB ||| ||| |||
DB 1057 GACCCGACCTTAAAGAAAGATAGTGTGATGCTCATGAGCAAGCAGATTTTTCGCCATA 1116
OY 184 LeuPro-----
DB ||| ||| |||
DB 1117 CTACCTCGATTGAAGCAGCTGGTGCATTAATATTATCCGGAACCAAGCTGCATGACC 1176
OY 186 AsnGlyAsnHisAlaLysTyrThrGlyCysValAlaIlePheAlaAla---- 200
DB 1177 GATATATAGCTATGCGCGCAAAATGATATTTCGCGCGGATTCGCATACCGATGTGCTG 1236
OY 201 -----GlnIleMetAlaGlyTyrHisSerTyrProLeuGln---GlyGlu 213
DB 1237 GGTATCTGGGCAAAATGATATTATTTATGCGCAAACTACTGGAAGCTAAAGATGTCTC 1296
OY 214 GlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGly-----Thr 230
DB 1297 AACAACTACGTTAGTCCGCTTACAAAGCTTATGCGCAATATGCGGAAAAAAGCTTAC 1356
OY 231 PheGly-----GluMetTyrAspTyrPheAsnMetProGlyAsnPro 244
DB 1357 TTCGTGATACCAAGTGTATGCGCAAACTCGATATTTGTCAATAGCTCGTCATGCT 1416
OY 245 AspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaIlePheLeuMetArgPheVal 264
DB ||| ||| |||
DB 1417 TCTGTACGAATGATCGACGCAAAAGAACTGCAT---CTCGTGTCAATGAATAAAGCAT 1473
OY 265 SerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyTyrTyr---SerValTyr 283
DB ||| ||| |||
DB 1474 GACAGCCCAATTCACGCCCCAATTTGATCTTCGCGCGCAAGACTTACATTTCCGCTAA 1533
OY 284 ValIleGlyAlaLeuArgAsnAsnPheAlaGlyTyrLysArgSerLeuGlnLeuHisVal--- 302
DB 1534 GTATGGGGTTCATAAACACGCTGCGCAAAATTAAGAAAGCAGCCCAATACGCCAATT 1593
OY 303 -----ArgAlaLeuTyrThrSerGlnGluTyrHisAspMetIleArgGlyLeu 319
DB 1594 TCAGGACACCGTTTACTTATACCTTATACCGTACCG-----CCTTTC 1629
OY 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIle-----GlyHis 336
DB ||| ||| |||
DB 1630 ACGCATATACATATTGCTGCTGACTACTGCGCAATAGCACAGCTCTCCAGGAAAGCTCGAA 1689
OY 337 AlaPheValCysAspGlyTyrAlaSerAspGlyTyrPheHisPheAsnTyr----- 353
DB 1690 AGCTTTAAGCTGAAAGCTGAGGCTGTGATGGGAAAGTCATTATCTGGATGCTTCC 1749
OY 354 -----GlyTyrPgly----- 356
DB 1750 AGCGAGTGTAGATACAGCGTACAGCGGCAACAGATGAAACGCCCTTCACTGCT 1809
OY 357 -----GlyValSerAsnGly----- 361
DB 1810 GTAGCATCCAACTTGACCGAAACGCTTATACGGAATACAAAGCTACAAACGCTTCA 1869
OY 362 -----PheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGluGly 1930
DB ||| ||| |||
DB 1870 TACTATTACAAAGTAAAC-----GCCAAACCAATTAAGGATGAGCAATCAATAT 1923
OY 381 GlyPheThrIleTyrGlnGluIleIleThrGlyIleGluProAlaLysThrProAlaGlu 400
DB 1924 TTGAAGAGCGTTCCGAAGATGCGCTGAACCGT-----CCGCTCCCTATGAAGCCGAA 1977
OY 401 AlaGlyTyrAspAlaLeuProIleLeuAlaLeuLysAsp-----IleGluAlaGluTyr 418
DB ||| ||| |||
DB 1978 GAAGCAGC-----CTGAAGGAACCAATTGTGGAATCCAGCGG 2016

```

```

OY 419 LysSerGluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlnGlnSer 438
DB ||| ||| |||
DB 2017 ACCGCTACTCCGGTCTGCTGTTATGACGAATTTCCAGAAATCCACAGGATCTCTGAC 2076
OY 439 -----AsnLeuAspLeuGlyTyrArg----- 445
DB 2077 ATGACATTCAGCTCCACAGCGCAGCTTGTCAATCTTACAAATGCGCTACCGTTCTCC 2136
OY 446 -----LeuAsn---LysAlaAspGlyGluValIleGlu 455
DB 2137 CATGATGACAAACGCCCAATTTCTCATTAAGCGCAAAAGGTTTGGGAACTCTGCT 2196
OY 456 ValLysThrSerSerIleAsnIleSerTyrPyrGlyTyrGlyGlnHisProGluSerPhe 475
DB 2197 AAGMAAACGCTGATTT-----
OY 476 SerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuTyrArgArg 495
DB 2224 TCCGAGAGCCAGAGTCTGTTGTAATGACAGCGCGCAATGCAATGATGATTTT-----GAA 2274
OY 496 ThrGlyThrGlnGlnTyrPglProValArgHisAlaGlnGlyIleTyrValAsnSerIle 515
DB 2275 ACAGCTGG---GGCTGTACGATATC-----GACTACGTCAGACTGAG 2316
OY 516 LysValAsnThrThrAspProAsnAsnValValIleThrValAspAsnAsnGlnGlyLys 535
DB 2317 CCTCGCCGTCAGCGCCCACTCATGCGGTAAACCAAAACGCTTACCAAT----- 2364
OY 536 LeuSerIleValIleProAsnSerPheValAlaAsp-----LeuAsn 548
DB 2365 -----CCGATGCGAGCGTAAAGCAAAAGCAATTCATGATCTGCTGAT 2412
OY 549 SerTyrGlnHisSerThrIleThrValGlnPheAsnSerAspSerProAspGluIleArg 568
DB ||| ||| |||
DB 2413 CAATACGGAAGAATATGCTCTCTGTCAGAGAAATTAAGAAATGATGCTTCA 2472
OY 569 ThrProValAlaPheAlaLeuSerThrGlyAlaThrAlaAspArgValIleSerLeuGly 588
DB 2473 GCCATGTA-----GTTAAAGCCGCGCATTCGAGCGCTGACCTGATGCACTATTCG 2526
OY 589 TyrValMetAlaGlu-----ValProGlyLysSerSerAsnTyrProValIleTyrSer 606
DB ||| ||| |||
DB 2527 CCAGCAGAGCGGAACAGCTTATGCTTACCCACAGAGCAAGCAAAAGCGATTCGATG 2586
OY 607 LysAsp-----ValLeuThrLeuSerGlyLysPyrThrLeuTyrTyr----- 621
DB 2587 AAGCAGGGGGGATCGTTACCTTTGCA-----TGCACCTGGAAACGA 2628
OY 622 ---ArgPheSerIleAsnAsnGln---LysAspGluTyrLysIleGlySerValSer 639
DB ||| ||| |||
DB 2629 CCGAAGTGTGATTCGATACGAGGGAAGAAATGCTGAGAGGCTTCTATGCCATTC 2688
OY 640 ValLysThrProThrGlnTyrThrHisProLeuPheGluValGlyHisAsnGlnThrSer 659
DB ||| ||| |||
DB 2689 ACCACATTCATTAAGATATGCG-----ATGATCATCCAGAGTCCGAA 2733
OY 660 ThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeu 679
DB ||| ||| |||
DB 2734 GATTATTAATTTACTTATTCGACATCGCATGATGATGAGG---CAATTGAACAACTTG 2790
OY 680 GlyLeuProPheAspGlyGluLeuValValPheArgGlnThrGlnSerSerGly 699
DB ||| ||| |||
DB 2791 CAG-----GATGCGAAGGTTCTGCTCTCTGCTTCCCTTGCAAGAGGGAAGCA 2841
OY 700 Ser-----LeuTyrAlaIleGlnGluThrValHisIleLysGln----- 712
DB 2842 AAATGTTCTGTGGGCGCAAGAGTCTGAGGCTGTTAAAGAGCTATATTATTTAATG 2901
OY 712 ----- 712
DB 2902 CACGACGTTGAGGAATGCAACAATTGAAACAATTCATTTGGTATGAAATCTGTT 2961
OY 713 -----GlyGluThrPheVal----- 717

```

Dd		2962	GCTCCGATGTGGTATCCCGAGACAGATATTGGATTTTGACSTTTGACTTTATCCG	3021
Oy		718	-----TCTGysProVal1-----	722
Dd		3022	CAGCAGAGTAGTATCACGCCCSAAATTTCAAAATACAGAACACTTGTTGCATTGGGCAAG	3081
Oy		723	-----GluLlyProIleProAsp-----	Gly 729
Dd		3082	GACAAAAAGTAGTGGCCCATGAGCGGMAAATGAGACCGGCGGACCCSGATTTGATGAG	3141
Oy		730	SeTYrAtgAlatPhLeuHisAlaPheValAsnGlnGlnGlnLeuTyLeuYsgLy	749
Dd		3142	CGSgATGCAAGCTCATTTGGAGCTGGTTCGCTACATGSgTAATGAGATTTTGTATGAGACGSC	3201
Oy		750	LysAlaGAsn-----TYrThValLysIleValAsnGlyTrhAlaValGlutAlaLe	766
Dd		3202	AACAACAAAACAGCCTTGAGCATCTGMAAAAGATGTATATCATCCGACAGCTCATTAAGCTG	3261
Oy		767	GLUeSerSerGluGluIleArgValPheProAsnProAlaAlaGAspTyValGluIleSer	786
Dd		3262	GATAGAG-----CTCCCAACGAACTTAAACAGCTATGCGATTACTGAG	3303
Oy		787	AlAProCySlleProGIngluThrSerIleIleLeuPheAspLeuSerGlyLysIleVal	806
Dd		3304	CAGCCGTCGCGACGGGACGCTTCACGCTGAACCTCGGGGTGAACGGCGMAAGTATCG	3363
Oy		807	MelLysAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArg-----LeuPro	824
Dd		3364	CTAAGCTGACAGACATCGCGMAATGCGMAAAGTATGMAAGTGAAGCGTTGCGACGACTGA	3423
Oy		825	AsnGlyAlaTyrr 828	
Dd		3424	AACGGCGCGTTC 3435	
 RESULT 19 AAK12949				
ID	AAK12949	standard; DNA; 9047 BP.		
XX AC	AAK12949;			
XX DT	19-MAR-1999	(first entry)		
XX DE	Enterococcus faecalis genome contig SEQ ID NO:12.			
KW RX	Enterococcus faecalis; confliq; detection; Enterococcal infection; vaccine; attenuation; computer readable medium; ds.			
Enterococcus faecalis.				
W09850555-A2.				
PD XX	12-NOV-1998.			
XX PE	04-MAY-1998; 98WO-US08985.			
XX PR	14-NOV-1997; 97US-0066009.			
PR	06-MAY-1997; 97US-0044031.			
PR	16-MAY-1997; 97US-0046655.			
XX PA	(HUMA-) HUMAN GENOME SCI. INC.			
XX PI	Barash SC, Dillon PJ, Kunsch CA;			
XX DR	WPI; 1999-045171/04.			
XX PT	New isolated Enterococcus faecalis polynucleotides and polypeptides			
PT	- used to develop products for the detection of Enterococcus and for			
PT	use in vaccines for prevention or attenuation of Enterococcus			
XX Infection.				
XX CS	Claim 1; Page 305-310; 2084pp; English.			
XX				

	CC	A computer readable medium has been developed which has recorded on it
	CC	982 nucleotide sequences isolated from the Enterococcus faecalis genome.
	CC	AAK12938 to AAK13191 represent these nucleotide sequences which are
	CC	primary nucleotide sequences, also known as contigs. The computer-based
	CC	system can identify fragments of the Enterococcus faecalis genome with
	CC	commercial importance. The products can be used to detect the presence
	CC	of Enterococcus faecalis in samples. They can also be used for
	CC	diagnosing Enterococcal infection in an animal and monitoring
	CC	progression of disease, and for identifying agents which can be used to
	CC	modulate the growth or pathogenicity of Enterococcus faecalis, or
	CC	another related organism, in vivo or in vitro. In particular the
	CC	polypeptides encoded by the Enterococcus faecalis nucleotide sequences
	CC	can be used in vaccines to prevent or attenuate an Enterococcal
	CC	infection.
	XX	
	SQ	Sequence 9047 BP; 3107 A; 1542 C; 1841 G; 2551 T; 6 other;
		Alignment Scores:
	Pred. No.:	0.0631 Length: 9047
	Score:	148.50 Matches: 182
	Percent Similarity:	34.59% Conservative: 120
	Best Local Similarity:	20.85% Mismatches: 320
	Query Match:	3.35% Indels: 252
	DB:	Gaps: 44
US-10-030-330-1 (1-843) x AAK12949 (1-9047)		
OY	143	glugluileleuatgthrGluGluyAAlProLaGlaValInIsAlaleMetaSPasngly 162
Db	1074	GAAGAAAGCATATCAGACGGACTTTCAGCGACGCCAC--CATGTCTGCTC----- 112L
OY	163	HISPhelaIAsnAPRProMetaIgtTPraSnglncltYUProTPraSnlpLySGlUPRo 182
Db	1122	-----AATCGCGAAGTACCCTGGAAATAAC---CAATTG 115L
OY	183	LeuleuPRAoNGLIyASnHIsAlAtytThrClcySuVaLAAlatPrAlalaGllle 202
Db	1152	TTCGATCCAAAAGGAAT-----GGCACGGCGGCACAAAGATT 119O
OY	203	--MetArgTYHISserTrpProLeuGlnGluGluSerPheASPtyHIsAlaGly 221
Db	1191	GTCAATTGATTTTTTGATACGAATGACAAAGAAAGAACATTTTGAACAACACGGCGT 125O
OY	222	SerleuVal-----GLyAnTPSer--glyThrHe----- 231
Db	1251	CSGATTATTACGGGTGGCGACSTACACAGAGGACTTTCAAACGTCGAATCTCAGG 131O
OY	232	-----GlyGIumetCytsprTriLeAsnMet-----ProGIyAsn 243
Db	1311	GCCGTGCAGTTAGTGAACCAAAGAATATGGGATTTTCCAATTTTAAGAASTGACAAATACA 137O
OY	244	PROASrleuAsPRASnleThrgInSerCIvalAsPlatryAlatHrleuMeTarGasP 263
Db	1371	CCAGAGCTAGTAATCTC-----GAAGCTTTAAGAAACCAAGGAGATCT 141L5
OY	264	VAlSerAlaseValSerMetSerPheTyGluAsnGlySerGlyThyrSerValTyg 283
Db	1416	CMAAGAGCGGGGATATTAGTACSTAGATATGACAAATGCAACAGTTCGAATAGTCTC-- 147Z
OY	284	VAlValGIyAlaleuArqAsnPrPheArGTyGlyUsArGSerLeuGlnleuHIsValArG 303
Db	1473	-----ACGAATAAGGAATAACTATCTTCATGAT----- 1499
OY	304	AlAleuTyTHrIserGlnGlyThrHIsArmetIleArGlyGylueAlaseGlyArG 323
Db	1500	-----GTTCACAAACAGATATATTTAATATATAGTGGGAATCTACCAACCAAGCA 155O
OY	324	ProVALtyUTyTAlaGlyAsnAsnGlsertIeGlyHIsAlaPheValIcysARPGlyTyg 343
Db	1551	-----TTAAACCGGTAGTGGAAATATATGTTTAAAGAAAGTTGG 1569
OY	344	AlaseARPGlyThrPheHIsPheAsnTPrclyTTPcgllyelAyAl----- 358

```

Db 1590 TTGTGAGAAAT-----GGCAATGATGTAACGGCAGCTAAT 1625
Oy 359 -----SeranSglyPheTyrLysLeuThrLeuLeuSerProthrSerLeuSly 374
Db 1626 GATCATCCGACACACGCTGGATTATATGATTGTAAGAAATATGATTACACA 1685
Oy 375 lIeGlyGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 384
Db 1686 GTGGGACTGTAAGGAATTCGACAAAGAGCTGATACCTCAAGAGCAACCTTTC 1745
Oy 385 TyrGln-----GluLeuLeuThrGlyLeu 392
Db 1746 TCAACAAACGCTTCTCAATATACATTCGTAATACATATTAAGATATCTCTCGGC 1805
Oy 393 GluProAlaLeuThrProAlaGluAlaGlyThrAspAlaLeuProAlaLeuAlaLeuS 412
Db 1806 GGGATTGCTCTGTCAAGACCAAAAGTGGCGGTGCTGGAAGAAATTTGCTGCCA 1865
Oy 413 AspIleGluAlaGluTyrLysSerGlyLeuAsnValGlyTyrSerIleTyrAsn 432
Db 1866 ATGTGCATGCCGATTATGATGACACAAATATACGGCGGCTTGGCAATGCTGTAGAT 1925
Oy 433 ThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 452
Db 1926 GACGACCTGCTCCATATATACGAAGATATACGAATCAATATGATTAACATGCTGA 1985
Oy 453 ValIleGluValLysThrSerSerIleAsnIleSerTrp---TyrGlyTyrGlyGly 471
Db 1986 GCTTATGATGTGACATGACATGACAAACATTTATATATACATATACATATACAT 2045
Oy 472 ProGlySer-----PheSerLeuAlaProAsnGlnLeuSerGlnGlyLeu---AsnThr 489
Db 2046 AATACTGGCGGCTTTTATGCTTATGCTGATCAAAAAAATCGGTTATGCTGCTCAAT 2105
Oy 489 ethrLeuLeuTyrArgArgThrGlyThr---GluGlnTrpGluProValArgHisAla 508
Db 2106 ATCTCACTATATGACGGCGGTGATACCGAGAACAGGTAAAGACATCCAGGC-GGACC 2164
Oy 508 nGlyGlyTyr-----ValAsnSerIleLeu 516
Db 2165 TCGTGGCTCAATTTATTAAGAACAAAGCATTTTCCATTTATGGTGAAAAATGATGGACC 2224
Oy 516 eValAsnThrThrAspProAsnAsnValValThrValAsp----- 530
Db 2225 GGCATATGCCACTATCCACATATATACGATTTATGTCGACGAGATTTTCAACATCAT 2284
Oy 531 -----AsnAsnGlyLysLeuSerIleValProAsnSerPh 543
Db 2285 ATTTGGTGAAGAAATAGTTCGACACACGCTGACAGTACCAATTTTACATATATAT 2344
Oy 543 eValAlaAspLeuAsnSerTyrGlyLysSerThrIleThrValGlnPheAsnSerAspSe 563
Db 2345 TTATA-----TACAAAGAAAGCTACCGGACAAATTAATCTTATCTAATATA 2389
Oy 563 rProAspGluIleArgThrProValAlaPheAlaLeu----- 575
Db 2390 TCCACAAATATGGACACAAACCAATGAAAGTGAATGTCGATTAATCCAGAAAAATATTT 2449
Oy 576 -----SerThrGlyAlaThrAlaAs 582
Db 2450 CAAAAATATATTTATTTGGCCAAAAAGAAATTTCTACAGAAAAAAGTGGTGCACATCTTGA 2509
Oy 582 pAspValIleSerLeuGlyTyrPValMetAlaGluValProGlyGlySerSerAsnTyrPr 602
Db 2510 AAAATATAGTATCAAGAGAAACATTTTGAAGAAACACAAATTAAGAAATACAGATTAACC 2569
Oy 602 oValValTrpSerLysAspValLeuThrLeuSerGlyGlyAspTyrThr----- 618
Db 2570 T-----GAAAGAGTGAAGAGTTGGCAGAGCAAAATTTACAGACATTTGAACC 2617
Oy 619 -----LeuTyrPyrArgPheSerIle 625
Db 2618 AACAAAGATATATGTTGATTGATTACTTCTAAAGAACGCTTACGTCAGCAGACAAAT 2677

```

```

Oy 625 eAsnAsnGlnLysAspGluTrpLysLysIleGlySer-----ValSerValLysThrPr 643
Db 2678 GTTCCGATTAATAAGAAACACTCCACCGCTTGTGTAAGCCTTGTGGAAGTGAACAGTCC 2737
Oy 643 oThrGlu-----TyrThrHisProLeuPheGluValGlyHisAsnG 657
Db 2738 GCCCGAAGACTCTTGTGAAATATAGCTTGAATAATTAAGTTTATGATATGTCGCCACA 2797
Oy 657 nThrSerThrTyrThrLeuAspMetAlaHis---AsnArgValLeuProAspPheThr 676
Db 2798 ACCATCCACAAATGAAAAAGATTCGTTATCAAAATCAAGGTTAGAAATTAAGTTTGGC 2857
Oy 676 uLysAsnLeu-----GlyLeuProPheAsnGlyLeuValVal-----ValPheArg 692
Db 2858 AACAGGTGTATTCCTAATTTCCACAGCAAGTGTGATTAACGTACAGAAAGTCTTAA 2917
Oy 692 gGlnThrGlnSerSerSerGlySer-----LeuTrpAlaAla-----G 705
Db 2918 TGAAGAAAGTGTGGCAGACGCAAAAAAGAAATTAAGTTACATGGAAGCAATCCCTCA 2977
Oy 705 nGluThrValHis-----IleLysGlnGlyGluThrPheValTyrLysProVa 721
Db 2978 AGAAAAATTAATATACAGCAGCAGCAGCAGCAGTGAAGTTCGACAGCAAGTTATGCTTACCGAT 3037
Oy 721 l-----ValGlyGlyProIleProAspGlySerTy 731
Db 3038 AGATGCTGTCAAGTATACAGCAGCAGCAGCAGTCTTTTGAAGGTGTAAGTGTGTAAGGAAAGCA 3097
Oy 731 rArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnGlnLeuTyrLeuLys----- 748
Db 3098 CACCGTTAACTCAAAACCA-----GCTCAAACTGCTATGCTTCAAAAGACGTGA 3145
Oy 749 -GlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaValAlaIleGlySe 768
Db 3146 TGGTAAACGACGATCTGTGCT---ATACAGCGGTACCGCT-----GCCATTAGTTC 3196
Oy 768 rSerGlnGluIleArgValPhePro-----AsnProAla----- 779
Db 3197 ACAGACGCGCTTAATAA---TATCCTTATGAGTAACATATACGGCAATTAACGATTAATA 3253
Oy 780 -----ArgAspTyrValGluIleSerAlaThr 788
Db 3254 ATTAATAATGCTACTCTGCGGGGTATATATGCGCGGATTAACGTTGAAATGATATACACA 3313
Oy 788 oCysIleProGlnGluThrSerIleLeuPheAspLeuSerGlyLysIleValMetTy 808
Db 3314 AGAATTGAAGAACTATCAATCACTAATAAAGTGCCTAATTTAGAACTAAAT---GTGATGCG 3370
Oy 808 sAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeuProAsnGlyAlaTy 828
Db 3371 TATATGATGCGTGAATGCTGCTGGAATATCAATATGATGAACGTTGAAAAAATACGACAGT 3430
Oy 828 rIleLeuLysValAspGlyTyrThrThrLysIleAsn 840
Db 3431 C-----CAAGTATGATTATGAACACAGATCTAAT 3461

```

RESULT 20

AA575934
ID AA575934 standard; cDNA; 9466 BP.

AC AA575934;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #11738.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.
PN WO2001/5067-A2.


```

Db 6282 GACCTGGAAGATGGACCTGCAAAAGTCTCTACTCTCCACCGCTGGGGCTTTATATC 6341
Oy 369 -----SerProthriserleuglyle 376
Db 6342 GTCTCCACCAAAATTCCTGACGACGACGTCCCGGAGCCCATTTACCGTGAAGATCAGT 6401
Oy 377 GlyGlyGlyLeuGlyPheThrIleTyrGlnGluIleIleThrGlyLeuProAlaLys 396
Db 6402 GGGGAGGGA-----AGATCAAAAGAGCATCACCCGACCAAGTCGGGCGCCG 6449
Oy 397 ThrProAlaGluAlaGlyThrAspAla-----LeuProIleLeuAlaLeu 411
Db 6450 TCCGGCCACATCGCGAGCATTTGTGACCTGAACCTGAATAATCCCAAAACCAACACC 6509
Oy 412 LysAspIleGluAlaGlyTyrLysSerGlySerGlyLeuAsnValGlyTyrSerIleTyr 431
Db 6510 AGGATATGTCGGCCACGTCACACCCCTCTGCG----- 6545
Oy 432 AsnThrGlyGlyGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGly 451
Db 6546 -----CGTGTGACTGAGGCA----- 6560
Oy 452 GluValIleGluValLysThrSerSerIleAsnIleSerTyrPyrGlyTyrGlyLysIle 471
Db 6561 GACATTGTGCCATG-----GGAGAGAAC 6584
Oy 472 ProGlySerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeu 491
Db 6585 TCACACTGGCGTCCGGTTGTGTGCCCCAGAGATG-----GGCGTGACACAGGTCACGTC 6638
Oy 492 LeuTyrArg----- 494
Db 6639 AACTACCGTGGGAGCAGCAGTCACCGGACCCCTTCAGTTCACCGTGGGCGACCTGGT 6698
Oy 495 -----ArgThrGlyThrGlnGlnTrpGluPro--- 503
Db 6699 GAAGGAGCGCCCAAGAGTGGCGGAGAGGCCCTGGCTGGAGAGAGAGAGAGAGCGGGA 6758
Oy 504 -----ValArgHisAlaGlnGlyTyrValAsnSerIle 515
Db 6759 GTCCAGCATGATTCAGCATTTGGACCCGGAGAGCGCCCTGA-----GGCCTC 6809
Oy 516 LysValAsnThrThrAspProAsnAsnValValValThrValAspAsnAsnGlu---Gly 534
Db 6810 TCCATGCGTGTGAGGCGCCAGTAAAGCCGAGATTAATGATGATCAATAAATAATGGG 6869
Oy 535 LysLeuSerIleValProAsnSerPheValAla---AspLeuAsnSerTyrGlnLysSer 553
Db 6870 TCGTGGGGTGA-----TCTTATATGCCCCAAGAGCCCTGTACTACGAG----- 6914
Oy 554 ThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThrProValAlaPhe 573
Db 6915 ---GTGTCCATCAAGTTCAAT-----GATGAGCACAATCCCGAAGACCCCTAC 6959
Oy 574 AlaLeuSerThrGlyAlaThrAlaAspAspValIleSerIleGlyTyrPheAlaGlu 593
Db 6960 CTGGTCCCGGTATCCACCCCTCCGAGAGCGCCGCGCTGATGAGACCTTCAG 7019
Oy 594 ValProGlyLysSerAsnTyrProValIleTrpSerLysAspValLeuThrLysSer 613
Db 7020 GAATCGGATTAATAAGTAAACCAAGCATCC----- 7052
Oy 614 GluGlyAspTyrThrLeuTyrTyrArgPheSerIleAsnAsnGlnLysAspGluTyrLys 633
Db 7053 -----TTTGCTATTAAGTTGATGCGCAAA----- 7079
Oy 634 LysIleGlySerValSerValLysThrProThrGlyTyrThrHisProLeuPheGluVal 653
Db 7080 -----GGCAGAATGATGCAAAAGTGCACAGC-----CCCTCTGGAGCGCTG 7121
Oy 654 GlyHisAsnGlnThrSerThrTyrThrLeuAspMetAlaHisAsnArgValLeuPro--- 672
Db 7122 GAGGAGTGCACATGCTGTAGCTGAGCCAGATAGTATGCTGTGCTTACCTTCAT 7181

```

```

Oy 673 AspPheThrLeuLysAsnLeuGlyLeuProPheAsnGlyGluLeuValVal----- 689
Db 7182 GAGATGTCGTCCACACCATCATGATGTCATTAAGGAGGACGAGTGTGGAGCCCC 7241
Oy 690 -----ValPheArgGlnThrGlnSerSerGlySerLeuThrAlaAlaGln 705
Db 7242 TTCAAAGTCGGGTGGGAGCCCTGGACACAGCGGGGAAACCCCTGCTGTCCCTAT 7301
Oy 706 GluThrValHisIleLysGlnGlyLysThr-----PheValLysPro 720
Db 7302 GGCAGG---GGAATCGAAGGGGGCACACAGGTATTCAGTCCGAATTTCTTTATACACC 7358
Oy 721 ValValGlyProIleProAspGlySerTyrArgAlaThrLeuHisAlaPheValAsn 740
Db 7359 ACCCGACAGGTGCA-----GGGACATTTATCCGTACACATCGAA 7397
Oy 741 Gly-GlnGlnGlnLeu-----TyrLeuLysGlyLysArgAsnTyrTh 754
Db 7398 GGCACCATCCAGGTTAAATGATGATGCCAGAAACACCTGAAGGGTACAAAGTCATTTAC 7457
Oy 754 rValLysIleValAsnGlyThrAlaValGluAlaIleGlySerSerGlyGluIleArgVa 774
Db 7458 ACCCCCATGGCTCGTGGTAACTACTGATCACTGATCAATACGGTGG----- 7506
Oy 774 ILeProAsnProAlaArgAspTyrValGlu----- 784
Db 7507 ---CCCAACACCATCGTGGGCAAGTCCCTTCAAGGCCAAGTGCACAGCCAGCT 7562
Oy 785 -IleSerAlaProCysIleProGlnGlnThrSerIleIleLeuPheAspLeuSerGlyLys 804
Db 7563 AGTTAGCCCTGC-CTCAGCCCAAGAGACCTCATCATCATCTGTGTGAGTCACTAGACCA 7621
Oy 804 sIleValMetLysAsnSer 810
Db 7622 GGTGCTCTCAAGAGACT 7640

```

Search completed: June 1, 2003, 22:41:03
 Job time: 633 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OW protein - nucleic search, using frame_plus.p2n model

Run on: June 2, 2003, 00:16:06 ; Search time 301 Seconds

(without alignments)
3776.318 Million cell updates/sec

Title: US-10-030-330-1

Sequence: 1 MKKSFILAVIMFGIAMOGH.....PNCAYILKVDGYTKRINIYH 843

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Dgapop 6.0	Dgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 845702 segs, 674182571 residues

Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame.p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO-SPOOL/US10030330/runat.23052003.181646.6824/app_query.fasta.1.1031
-DB=Published_Applications_NA -QMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100
-THR.MIN=0 -ALIGN=20 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10030330.ecgn.1.1.170.6runat.23052003.181646.6824
-NCPU=6 -ICPU=3 -NO.XLPHY -NO.MAP -LARGEQUERY -NEG.SCORES=0 -WAIT
-DSBLOCK=100 -LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10
-XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications_NA:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	403.5	9.1	1419	US-10-002-784A-23	Sequence 23, Appl
2	154	3.5	3878	US-09-741-669-111	Sequence 111, App
3	148.5	3.3	4059	US-09-784-554B-1	Sequence 1, Appl
4	148.5	3.3	9047	US-09-070-927A-12	Sequence 12, Appl

Result	Score	Query Match	Length	ID	Description
5	145	3.3	25165	US-10-114-170-39	Sequence 39, Appl
6	140	3.2	3106	US-08-837-459-21	Sequence 21, Appl
7	139.5	3.1	3695	US-10-114-170-18	Sequence 18, Appl
8	138.5	3.1	4605	US-10-080-505-10	Sequence 10, Appl
9	138.5	3.1	4822	US-10-080-505-14	Sequence 14, Appl
10	138.5	3.1	10455	US-10-160-758-4	Sequence 4, Appl
11	138	3.1	4765	US-10-090-624-5	Sequence 5, Appl
12	138	3.1	4828	US-10-080-505-16	Sequence 16, Appl
13	138	3.1	9249	US-09-848-616-169	Sequence 169, App
14	138	3.1	9299	US-09-848-616-168	Sequence 168, App
15	137.5	3.1	4268	US-10-114-170-93	Sequence 93, App
16	136.5	3.1	11065	US-10-160-758-1	Sequence 1, Appl
17	136.5	3.1	11073	US-10-160-758-2	Sequence 2, Appl
18	133.5	3.0	7712	US-09-974-300-2632	Sequence 2632, App
19	133	3.0	4557	US-10-114-170-206	Sequence 206, App
20	133	3.0	43360	US-10-114-170-261	Sequence 261, App
21	133	3.0	45345	US-10-114-170-261	Sequence 20, Appl
22	132.5	3.0	3131	US-08-837-459-20	Sequence 2739, Ap
23	131	3.0	8979	US-09-738-626-2739	Sequence 1, Appl
24	131	3.0	3309400	US-09-738-626-1	Sequence 1, Appl
25	127.5	2.9	3309400	US-09-738-626-1	Sequence 263, App
26	126.5	2.8	2637	US-09-912-020-204	Sequence 204, App
27	125	2.8	5010	US-09-917-800A-483	Sequence 164, App
28	125	2.8	5681	US-09-848-616-164	Sequence 3, Appl
29	125	2.8	4056	US-09-784-554B-3	Sequence 147, Appl
30	124	2.8	7198	US-10-114-170-147	Sequence 5, Appl
31	123.5	2.8	2604	US-09-839-894-5	Sequence 2, Appl
32	122.5	2.8	7239	US-09-839-894-27	Sequence 29, App
33	122.5	2.8	6702	US-09-769-787-209	Sequence 8615, Ap
34	122	2.7	7035	US-09-815-242-8615	Sequence 12, Appl
35	121	2.7	5245	US-10-080-505-12	Sequence 50, Appl
36	120.5	2.7	1791	US-10-213-090-50	Sequence 59, Appl
37	120	2.7	2870	US-10-121-032-59	Sequence 59, Appl
38	120	2.7	2870	US-10-093-037-59	Sequence 2, Appl
39	120	2.7	3169	US-10-131-591A-2	Sequence 404, App
40	120	2.7	3597	US-09-975-719-404	Sequence 8512, Ap
41	120	2.7	17388	US-09-815-242-8512	Sequence 1, Appl
42	120	2.7	42235	US-09-975-719-1	Sequence 7, Appl
43	120	2.7	4794	US-10-092-880-7	Sequence 200, App
44	119.5	2.7	4794	US-10-092-880-7	
45	119	2.7	6091	US-10-114-170-200	

ALIGNMENTS

RESULT 1
US-10-002-784A-23
; Sequence 23, Application US/10002784A
; Publication No. US200303644A1
; GENERAL INFORMATION:
; /33
; APPLICANT: Ulrich, Robert G.
; TITLE OF INVENTION: Bacterial Superantigen Vaccines
; FILE REFERENCE: 003/233/SAP
; CURRENT APPLICATION NUMBER: US/10/002,784A
; PRIOR FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 08/882,431; 09/144,776
; PRIOR FILING DATE: 97-06-25; 98-09-01
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 23
; LENGTH: 1419
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: mutant Spea/mutant Speb fusion
US-10-002-784A-23

Alignment Scores:
Pred. No.: 7.49e-33
Score: 403.50
Percent Similarity: 50.17%
Best Local Similarity: 34.65%

Length: 1419
Matches: 105
Conservative: 47
Mismatches: 114


```

1372 GGAACCCCAATATTAGGATGGCGGTAGTTCGATCGCAAAATATTACTTT----- 1425
OY 352 AsnTrpGlyTrpGlyValSerAsnGlyPheTrpLysLeuThrLeuLeuSerProThr 371
Db 1426 -----TTGGCGCATAAAGACAGCGGTATCGGCACACTGACACTGATGATGCGACA 1476
OY 372 Ser-----LeuGlyIleGlyGlyGlyIle----- 380
Db 1477 TCATCGTTCGATCTGGGTATCATATGCTGTAATTTGGTGGCGGTATCGTAAATGTC 1536
OY 381 -----GlyPheThrIleTrpGlnGlu-----IleIleThrGlyIleGluProAla 395
Db 1537 AGTATGGTGGCCACCCCTTAATTCACAGCGGTATGATTTATCGGAGAAATCCCGGT 1596
OY 396 LysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlu 415
Db 1597 AAGGGAGATGTAATTAATTTCAACGCGACAGCTCTGCAATTTAAACATCATCTACTAAC 1656
OY 416 AlaGluTrpLysSerSerGlySerGlyLeuAsnValGlyTrpSerIleTrpAsnThrGlyGlu 435
Db 1657 GCACATATG-----CTACAGCTCGGT-----GATTTAGGCGCGGATGA 1695
OY 436 GluGlnSerAsnLeuAspLeuGlyTrpArgLeuAsnLysAlaAspGlyValIleGlu 455
Db 1696 -----GluProVal-----CTGATATTACACCGCGAGTACTTAA 1725
OY 456 ValLysThrSerSerIle-----AsnIleSerTrpTrpGly 467
Db 1726 GCGCGTGTACACAGATAGCTCTCAATGACAAATAGCGGAGCGAGGTGATGG 1785
OY 468 TyrGlyGlnHisProGlnSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsn 487
Db 1786 CAGACTCTCTTTGTAACATCATATGTAAGTGG-----ACATCTGCTACGGGT 1839
OY 488 ThrIleThrLeuLeuTrpArgTrpThr-----GlyThrGlnGlnTrp----- 501
Db 1840 ACGTTAACCTCGACGAATACAGGTACGCTGATGTCGAAGTGGAAGATTACTTAGGT 1899
OY 502 -----GluProVal-----ArgHisAla 507
Db 1900 GTTTTGTAGCTGCTGTAGAACCGTTAAACATTTGCTGCTCACGCTGAGCGCGCGCA 1959
OY 508 GlnGlyGlyTrpVal-----AsnSerIleLysVal----- 517
Db 1960 GATCCCGATTTATCACCAATGCGACGAAAGTGAGTTGGCTTTGGCGAAGCGCTTTT 2019
OY 518 -----AsnThrThrAspProAsn-----AsnValValValThrVal 529
Db 2020 GTCTTAAATCATACCAATACAGATGATCCGCGTACAGGTCGATATGCTGATACAGGT 2079
OY 530 AspAsnAsnGlnGlyLysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSer 549
Db 2080 GACGATAAAGACGGAAAGTATC----- 2103
OY 550 TyrGlnHisSerThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThr 569
Db 2104 -----CATGATGACAGGCCATACGCTGTTCAATGACGGAATATTAAAGCGTAAACG 2157
OY 570 ProValAlaPheAlaLeuSerThrGlyAla-----ThrAlaAspValIleSerLeu 587
Db 2158 CTGGTCATGACGGCGCTCTCGACATTCGCTCTCATAGCGGAGATGAGGTTAAGCGGCATG 2217
OY 588 GlyTrpValMetAlaGluVal-----ProGlyGlySerSerAsnTrpProVal 603
Db 2218 GGG-----TCGAGTGAAGTACCAATTCGCAAAACCCCGT----- 2250
OY 604 ValTrpSerLysAspValLeu-----ThrLeuSerGlyGlyAspTrpThrLeuTrpTrp 621
Db 2251 -----ACGCTCGACATTCGTCATCAACGAAACAGTGCAGAGATTTACACGCTGACCAAT 2304
OY 622 ArgPheSer-----IleAsnAsnGlnLysAspGluTrpLysIleGlySer 637

```

```

Db 2305 GCGCTCAAAAGGAGTGGCTTGATGCGAGTGCAGCTGTATCCCGACAAAGATGTTGGC 2364
OY 638 ValSerValLysThrProThrGluTrpThrHisProLeuPheGluValAlaHisAsnGln 657
Db 2365 TTATACCCATGACAAAGGAGCTGAATTCGCC-----GGTGTCCCAACTGAA 2412
OY 658 ThrSerThrTrpThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLys 677
Db 2413 GACAGTACCTTCATCTGGACCGCAGCAACACCGCTCGCTGCTTCTACCGCATGTTGCAG 2472
OY 678 AsnLeuGlyLeuProPheAsnGlyLeuValValAlaPheArgGlnThrGlnSerSer 697
Db 2473 TCT-----GACAGTGAATAATACCATCGGTAAATTTGGAGCAATCC 2517
OY 698 SerGlySerLeuThrAlaAlaGlnGluThrValHisIleLysGlnGlyGluThrPheVal 717
Db 2518 ATTGGTGGACTG-----GCCATGATATGAGATGACCAATCAT 2553
OY 718 Tyr-----LysProValValGlnGlyProIleProAspGlySerTrp 731
Db 2554 TTGATACGGAATATTCCTCTCGACGCTCGCGGAGGATATATC----- 2598
OY 732 ArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnLeuTrpLeuLysGlyLysArg 751
Db 2599 -----ACGCTGATACGCTGTTGTCGCGCGGCGGTGATACACCTGGAAGGC---CGT 2649
OY 752 AsnTrpThrValLysIleValAsnGlyThrAlaValAlaGluAlaIleGluSerSerGlyGlu 771
Db 2650 AACTATACG-----GTAAACGGAGCGGCGAGCTGCTTATCAGC----- 2688
OY 772 IleArgValPheProAsnProAlaArgAsp 781
Db 2689 -----GTGCTTAAACGTGGAATGAT 2709

RESULT 3
US-09-784-554B-1
; Sequence 1, Application US/09784554B
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schmoor, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schuelen, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGUCANASES
; FILE REFERENCE: 10017, 200-US
; CURRENT APPLICATION NUMBER: US/09/784, 554B
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4059
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-1

Alignment Scores:
Pred. No.: 3 56e-05 length: 4059
Score: 148.50 Matches: 176
Percent Similarity: 32.10% Conserved: 127
Best Local Similarity: 18.64% Mismatches: 334
Query Match: 3.35% Indels: 307
DB: 9 Gaps: 47

US-10-030-330-1 (1-843) x US-09-784-554B-1 (1-4059)
OY 112 AspMetAspSerMetProAspAsnLeuArgMetTrpLeuGlnIleTrpAspGlnIle 131
Db 844 GATTCGAGCTCTGTAAAGCAATATATACCTGTGCTGACATTAATACCTGATCAATG 903
OY 132 GlyLeu----- 133
Db 904 GCGCTTACCTGCAAGTCGACAGCAAGACATTCGATGATTCGACGCTACACTGAT 963
OY 134 -----IleLeuSerGlyLysAlaGlnLeuAsnGlnGluIleLeuArgThrGlnGlyVal 151

```

```

Db 964 CCCGAGCGATGGGGGAGGAGCATACGAATACGAATGAGGTAGGCAATGACAAACGAAG 1023
Oy 152 ProAGluValHisAlaLeuMetAspAsnGlyHisPheAlaAsnAspProMetArgTrp 171
Db 1024 AAGCCAGAAATCCAGCA-----CCTGGCACTTTGTG 1056
Oy 172 AsnGlnGlyTrp-----ProTrpAsnAsnLys---GluProLeu 183
Db 1057 GACCCGACTATAGGAAGATAGTTGATCGCTCAATGGAACAGCAAGTTTGGCCANA 1116
Oy 184 LeuPro----- 185
Db 1117 CTACCTCGATTGACAGCACTGGTGATAAATATTATCCGGAGCAACAGCTGCCAATGACC 1176
Oy 186 -----AsnGlyAsnHisAlaTrpTrpGlyCysValAlaThrAlaAlaAla---- 200
Db 1177 GAGTATAGCTATGGGGGCAAAATGATATTTCCGGGGATTTGCCATGACCATGTGCTG 1236
Oy 201 -----GlnIleMetArgTrpHisSerTrpProLeuGln---GlyGlu 213
Db 1237 GGTATCTTGGCCAAAATGATGTTTATATGCAAACTACTGCACTAAAGATGGTGTC 1296
Oy 214 GlySerPheAspTrpHisAlaGlySerLeuValGlyAsnTrpSerGly-----Thr 230
Db 1297 AACAACTACGTTAGTGCCGCTTACAACTTTATCCGAATATGACGAAACAACTCTACT 1356
Oy 231 PheGly-----GluMetTrpAspTrpIleAsnMetProGlyAsnPro 244
Db 1357 TTCCGTGATACCAAGTGTAGTGCAGCAACATGCGATATGTCATATAGCTGGTCCATGCT 1416
Oy 245 AspLeuAspAsnLeuThrGlnSerGlnValAspAlaTrpAlaThrLeuMetArgAspVal 264
Db 1417 TCTGTACGAATGCATCCGACAAAGAACTGCAT---CTCGTTGTCAATGAATAAAGCAAG 1473
Oy 265 SerAlaSerValSerMetSerPheTrpGluAsnGlySerGlyThrTrp---SerValTrp 283
Db 1474 GACACGGCATTCGACGCCCAATTTGATCTTCCGGGGCAGAGCACTTACCTTCGGGTAAA 1533
Oy 284 ValValGlyAlaLeuArgHisAsnAspPheArgTrpLysArgSerLeuGlnLeuHisVal--- 302
Db 1534 GTATGGGGTTCGATAAACAGCTCGCAATTTAAAGAACAGCCCAATCACGCAAAAT 1593
Oy 303 -----ArgAlaLeuTrpThrSerGlnGluTrpHisAspMetIleArgGlyGluLeu 319
Db 1594 TCAGCAACCGTTTACTTATACCTACCTACCG-----CCTTTG 1629
Oy 320 AlaSerGlyArgProValTrpTrpAlaGlyAsnAsnGlnSerIle-----GlyHis 336
Db 1630 ACGGATATACATTTGTGTCTACTGCTACGTGCAATGACACGCTCTCCAGTGGAGGTCCTGAA 1689
Oy 337 AlaPheValCysAspGlyTrpAlaSerAspGlyThrPheHisPheAsnTrp----- 353
Db 1690 AGCTTTAAGCTGAAGAGCTGAGCTGTGTATGGGAAAGTCCATTATTCCTGGGATGCTTCC 1749
Oy 354 -----GlyTrpGly----- 356
Db 1750 AGCGAGTGTAGGATACAGCGCTACAGCGGCGCAACAGATGAAGAGCCCTTTCATCTGT 1809
Oy 357 -----GlyValSerAsnGly----- 361
Db 1810 GTAGCATCACTTACCGAAACGCTTATACAGGATACCTACGTCGCAACGCTACTCTCA 1869
Oy 362 ---PheTrpLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyCysGlyGluGly 380
Db 1870 TACTATTACAAAGTAAC-----GCCAAACCAATTAAGGATCGACGAATCCCAATAT 1923
Oy 381 GlyPheThrIleTrpGlnGluIleIleThrGlyIleGluProAlaLysThrProAlaGlu 400
Db 1924 TTGAAGAGGTTCCGAAGATGCTGTAAGCGT-----CCGCTCGCTATGAAGCGGA 1977
Oy 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAsp-----IleGluAlaGlyTrp 418

```

```

Db 1978 GAAGCAGC-----CTGAAGGCAACATTTGTGAAATCCAGCGG 2016
Oy 419 LysSerGluSerGlyLeuAsnValGlyTrpSerIleTrpAsnThrGlyGluGlnSer 438
Db 2017 ACCGGCTACTCCGGTGGTGTGTATGACGAATTTCCAAATCCAGGGATTTCTGTACG 2076
Oy 439 -----AsnLeuAspLeuGlyTrpArg----- 445
Db 2077 ATGACGATTCAAGCTCCACGAGCGCTGTGCAATCTTACAAATCGGCTACCGCTCTCT 2136
Oy 446 -----LeuAsp---LysAlaAspGlyGluValIleGlu 455
Db 2137 CATGATGACAAACGACCAATTTCTCATTAACGCGAAAGCGTTGGCGCACTGCTGCTT 2196
Oy 456 ValLysThrSerSerIleAsnIleSerTrpGlyGlyGlyGlnHisProGluSerPhe 475
Db 2197 AAGMAACGGCTGATTT-----AAAGAACT 2223
Oy 476 SerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTrpArg 495
Db 2224 TCCGGAGCGCAAGCTGCTGTGATGACGCGCGCAATACGATCAGTTT-----GAA 2274
Oy 496 ThrGlyThrGluGlnTrpGluProValArgHisAlaGlnGlyGlyTrpValAsnSerIle 515
Db 2275 ACAGGCTGG---GGCTGTACGATATC-----GACTACGTCAGACTGGAG 2316
Oy 516 LysValAsnThrThrAspProAsnAsnValValThrValAspAsnAsnGluGlyLys 535
Db 2317 CCTGGCGGTGACCGCCACCTCATGGGGAACCAAAAGCCTTACCAAT----- 2364
Oy 536 LeuSerIleValProAsnSerPheValAlaAsp-----LeuAsn 548
Db 2365 -----CCAAATCGCAGCGTAGAAGCAAAAGCATTGATGAACTATCTGTGAT 2412
Oy 549 SerTrpGluHisSerThrIleThrValGlnPheAsnSerAspSerProAspGluLeuArg 568
Db 2413 CAATACGGGACAAATATGCTCTGTGTCAGAGGAATTAACGAATTAATTTGCGTTCAA 2472
Oy 569 ThrProValAlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGly 588
Db 2473 GCCAATGTA-----GTAATAAGCGCGGATTCGACGCGCTGACCTGATCGACATTCG 2526
Oy 589 TrpValMetAlaGlu-----ValProGlyGlySerSerAsnTrpProValValTrpSer 606
Db 2527 CCAGCAGAGCGGAACACGCTTGTATGTCACAGAGCGAAGAGCGGATTCATGGGAT 2586
Oy 607 LysAsp-----ValLeuThrLeuSerGluGlyAspTrpThrLeuTrpTrp----- 621
Db 2587 AAGCAAGGGGGATGCTTACCTTGCA-----TGGCACTGGAAACGA 2628
Oy 622 ---ArgPheSerIleAsnAsnGln---LysAspGluTrpLysLysIleGlySerValSer 639
Db 2629 CGGAAGAGCTGATGATACGAGGGAAGAAAGATGAGAGGCTGTATCCGATTTCA 2688
Oy 640 ValLysThrProThrGluTrpThrHisProLeuPheGluValGlyHisAsnGlnThrSer 659
Db 2689 ACCAATTCGATATGAAATATCG-----ATGAATCATCCACAGTCGGA 2733
Oy 660 ThrTrpThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeu 679
Db 2734 GATTATTAATTACTTATTCGACATCCATGTGATTCGACGG---CAATTGAAGAATTTG 2790
Oy 680 GlyLeuProPheAsnGlyGluLeuValValPheArgGlnThrGlnSerSerSerGly 699
Db 2791 CAG-----GATGCGAAGTCTCTGTCGTCGTCCTTGCACAAACGGAAGCA 2841
Oy 700 Ser-----LeuTrpAlaAlaGlnGluThrValHisIleLysGln----- 712
Db 2842 AATGTTCTGTGGGGCGCAAGAGTCCTGAGCCTGTAAAGAACTGTATTTAATG 2901
Oy 712 ----- 712
Db 2902 CACGACGTTTGACGATGTGCAAAATTTGAATTTGGGTATGGAATTCGT 2961

```


QY 713 -----GlygluThrPheVal----- 717
 Db 2962 GCTCGGATGTGATCCGGAGACAGATGTGATATTGAGCTTGACTTATCCG 3021
 QY 718 -----TyrLysProValVal----- 722
 Db 3022 CAGCAGGTGATTAACGCCGCCAATTTCAAAATACGAACCTTGTTCATTGGCGA 3081
 QY 723 -----GluGlyProIleProAsp-----Gly 729
 Db 3082 GACAAAAAGCTAGTTCATGAGCAAAATGAGACCGATCCGACCCCTGATTGTATGANG 3141
 QY 730 SerTyrArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnLeuTyrIleuLysGly 749
 Db 3142 GCGTATCAGCTCATTTGAGCTGTGCTCATCATGTATGAGATTTTGTGAGACGCGC 3201
 QY 750 LysArgAsn-----TyrThrValLysIleValAsnGlyThrAlaValGluAlaIle 766
 Db 3202 AAACAAAACAGCCTTGAGCATCTGAAAAAGTGTATATCATCCGACGCTCATTCGCTG 3261
 QY 767 GluSerSerGluGluIleArgValPheProAsnProAlaArgAspTyrValGluIleSer 786
 Db 3262 GATGAG-----CTCCCAACGAATTAAAGATGATGATTAAGTACTGAG 3303
 QY 787 AlaProCysIleProGlnGluThrSerIleIleLeuPheAspLeuSerGlyLysIleVal 806
 Db 3304 CAGCGCTCCGATCCGGAGCTTCACGCTGCAACGCTGCGGTGAACGCGCAAGATGCG 3363
 QY 807 MetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArg-----LeuPro 824
 Db 3364 CTAGAGCTGACAGCATCGCGCAATGCGAAAGCTATGAAGTGAAGCGTTGACGACTGAA 3423
 QY 825 AsnGlyAlaTyr 828
 Db 3424 AACGCGCGCTC 3435

RESULT 4

US-09-070-927A-12

Sequence 12, Application US/09070927A

Patent No. US20020120116A1

GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch

Patrick J. Dillon

Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 982

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070, 927A

FILING DATE: 04-May-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB369
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9047 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 12:
 US-09-070-927A-12
 Alignment Scores:
 Pred. No.: 0.000121 Length: 9047
 Score: 148.50 Matches: 182
 Percent Similarity: 34.59 Conservative: 120
 Best Local Similarity: 20.85 Mismatches: 320
 Query Match: 3.358 Indels: 252
 DB: 10 Gaps: 44

US-10-030-330-1 (1-843) x US-09-070-927A-12 (1-9047)

QY 143 GluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMetAspAsnGly 162
 Db 1074 GAAAAAGTAAATCAGCGAGATTTCAGCCAGCGGAC---CATGTTGCTG----- 1121
 QY 163 HisPheAlaAsnAspPrometArgTyrPasnGlnGlyTyrProTyrPasnAsnLysGluPro 182
 Db 1122 -----AATGGCGCAAAAGTACCTGGAAATAC---CAATG 1151
 QY 183 LeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAlaThrAlaAlaGlnIle 202
 Db 1152 TTGCATCCAAAAGGAAAT-----GGGACACGCGCACAAAAGATT 1190
 QY 203 ---MetArgTyrHisSerTyrProLeuGlnGlyGluGlySerPheAspTyrHisAlaGly 221
 Db 1191 GTCATTTGATTTTATATATAGATGACAAAGGAAAGCAATTTTAAACACACGCGCT 1250
 QY 222 SerLeuVal-----GlyAsnTyrSer---GlyThrPhe----- 231
 Db 1251 CCGATTATTAAAGCGTGGTGCACCTACACACAGGAGCTTTCAACGTGCAATCTCAGG 1310
 QY 232 -----GlyGluMetTyrAspTyrIleAsnMet-----ProGlyAsn 243
 Db 1311 GCCGTGCAGTTAGTGAACCAAGAAATTTGGAAATTTCAAAATTTAAAGTGAACAAATACA 1370
 QY 244 ProAspLeuAspAsnLeuThrGlnSerGlnValAlaPheValThrLeuMetArgAsp 263
 Db 1371 CCAGAGCTGATTAATCTC-----GAAGGCTATTAAGAAACAGGAGATGCT 1415
 QY 264 ValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyrSerValTyr 283
 Db 1416 CAAGAGCGCGGATATTAGTACTAGATATGAACAAATGCAACGTTCAATAGTGC--- 1472
 QY 284 ValValGlyAlaLeuArgAsnAsnPheAlaGlyTyrLysArgSerLeuGlnLeuHisValArg 303
 Db 1473 -----ACGATTAAGAAATTAATCTGTTCAAGAT----- 1499
 QY 304 AlaLeuTyrThrSerGlnGluThrPheAspMetIleArgGlyGluLeuAlaSerGlyArg 323
 Db 1500 -----GTTCAACACAGAAATATTATTTAAATCTAAGGAGGAATCTGCACCAACGA 1550
 QY 324 ProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyr 343
 Db 1551 -----TTAAAAGCGGTAGTGGATTAATTTTAAAGAACTGG 1589
 QY 344 AlaSerAspGlyThrPheHisPheAsnTyrPglTyrGlyGlyVal----- 358
 Db 1590 TTTCATGAGAAAT-----GGCAAGCTATTAACGCGACGTAAT 1625
 QY 359 -----SerAsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGly 374

```

Db      1626 GATCATCGAACCAACGCTGATTTAATGATATTTGATGATAAATAATGATGATCAACGA 1685
OY      375 11eglygluglylle-----Glypethrille 384
Db      1686 GTGGAGCTTGGAAGAAATTCGAACAAAGCTGATTCAGATACCCAGAGGCAACACTTTC 1745
OY      385 TyrGln-----Glutellethrlylle 392
Db      1746 TACAAAGCTTTTATATATGATTCGATTCGATTAATGAAAGATATTCCTGGCGAC 1805
OY      393 GluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuProIleuAlaLeuLys 412
Db      1806 GGGATGCTCTTCACAGCAAGCAAAAGTGGCTGTCGAGAAAGAAATTCCTGCGCA 1865
OY      413 AspIleGluAlaGlyThrLysSerGlyLysLeuAsnValGlyTyrSerIleTyrAsn 432
Db      1866 ATGTGCAGATGCCATATGTGTACACAAATACGGGGCGTTGGCAATGCTGTAGAT 1925
OY      433 ThrGlyGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlu 452
Db      1926 GACGAGCTTCCATATATACAGATATACGGAATCAAAATATGATTTACGATGCTGAA 1985
OY      453 ValIleGluValLysThrSerSerIleAsnIleSerTrp---TyrGlyTyrGlyGlnHis 471
Db      1986 GCTTATGATGTGACATGACATGCAAAACATGTTATATCATATATACATTCACATCAT 2045
OY      472 ProGluSer-----PheSerLeuAlaProAsnGlnLeuSerGlnGlyLe---AsnThrIle 489
Db      2046 AATACGTGGCGCTTTTATGCTTATGCTGATCAAAAATTCGGTTATTCGGTCAAT 2105
OY      489 eThrLeuLeuTyrArgThrGlyThr---GluGlnTrpGluProValArgHisAlaGlu 508
Db      2106 ATCCACTATATGACGGCGGTGTACCGAGAACGATTAAGACATCCAGGC---GGAGC 2164
OY      508 nGlyGlyTyr---ValAsnSerIleLys 516
Db      2165 TGGTGGCTACATTTATAAGAACAAAGCATTTTCATATGCGTGAATAAATGATGAGC 2224
OY      516 sValAsnThrThrAspProAsnAsnValValValThrValAsp-----530
Db      2225 GGCATGCCACATATCCACATATATACGATTTATGTCGCGCGATATTCACATCAT 2284
OY      531 -----AsnAsnGlnGlyLysLeuSerIleValProAsnSerPh 543
Db      2285 ATTGGTGAGCAATAGTTCGACACAGGTGACAGTACCCAAATTTTACATATATAT 2344
OY      543 eValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPheAsnSerAsp 563
Db      2345 TTTA-----TACAAAGAGGTACCGGACAAATTAATTTCTTATCTAATTA 2389
OY      563 rProAspGluIleArgThrProValAlaPheAlaLeu-----575
Db      2390 TCCAAACAAATGGACACACCAATTGAAAGTGAATGCTGATTCAGAAAATATAT 2449
OY      576 -----SerThrGlyAlaThrAlaAs 582
Db      2450 CAAAATATATGTTATTTGCCAAAAGAAATTCGTACAGAAAAGGCGTCACTGTGA 2509
OY      582 pAspValIleSerLeuGlyTyrPheAlaGluValProGlyLysSerAsnTyrPr 602
Db      2510 AAATATAGTATCAAGCAACATTTTGAAGAACCAATTAAGAAATCACAGATTAACC 2569
OY      602 oValValTyrSerLysAspValLeuThrLeuSerGlnGlyAspTyrThr-----618
Db      2570 T-----GAAAAGTGAAGAGTGGCAGACGACAGCAATTCAGACATTTGAACC 2617
OY      619 -----LeuTyrPheSerIle 625
Db      2618 AACAAAGATATATGTTGTAATTTACTCTTAAGAACGCTTACGTCAGGACACAAAT 2677
OY      625 eAsnAsnGlnLysAspGluTyrLysLysIleGlySer-----ValSerValLysThrPr 643
Db      2678 GTTCGATTAAGAAAGAACTCCAGCGATTGTAAAGCGTGTGCGAAGTGAACGCTC 2737

```

```

OY      643 oThrGlu-----TyrThrHisProLeuPheGluValGlyHisAsnGlu 657
Db      2738 GGGCAAGACTCTTTGGAAATAGCTTGAAGAAATTAAGTTTATGATATTTGGCGCACAA 2797
OY      657 nThrSerThrTyrThrLeuAspMetAlaHis---AsnArgValLeuProAspPheThr 676
Db      2798 AGCATCCACAAATTTGAAGAAAGATATTCGTTATCAAAATATCAAGGTATGAAGATTA 2857
OY      676 uLysAsnLeu-----GlyLeuProPheAsnGlnGlyLeuValVal-----ValPheArg 692
Db      2858 AACAGGTGTATCTCTTAATTTCCAGACCAAGTTGAATTAACGATACAGAAAGTCTTAA 2917
OY      692 gGlnThrGlnSerSerSerGlySer-----LeuThrPalaIle-----Glu 705
Db      2918 TGAAGAAAGTGTGGCAGACAGCAAAAAGAAATTTAAGTTACAAATGGCAAGCAATCCCTCA 2977
OY      705 nGluThrValHis-----IleLysGlnGlyGluThrPheValTyrLysProVal 721
Db      2978 AGAAATAATTAATACAGACAGCACAGTTGAAGTTGACGACAAAGTTATGCGCTTACCGAT 3037
OY      721 1-----ValGluGlyProIleProAspGlySerTyr 731
Db      3038 AGATGCTGTCAAAGTACAGCAAGGCTCTTTTGGAGGTGAACTGTGGAAGAAAGAA 3097
OY      731 rArgAlaThrLeuHisAlaPheValAsnGlnGlnGlnLeuTyrLysLys-----748
Db      3098 CACGTTAACTTAACAA-----CCTCAAGCTGCTATGTTCAAAAGAGTGA 3145
OY      749 -GlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaValAlaIleGlu 768
Db      3146 TGGTAAACCGACATCTGCT---ATACGAGCGGTACCGCT---GCCATTAGTTC 3196
OY      768 rSerGlnGluIleArgValPhePro-----AsnProAla-----779
Db      3197 AGGAGAGCGCTTAAAT---TATCTTATGAGTAACTATACGGGCAATTAACGATTTAA 3253
OY      780 -----ArgAspTyrValGluIleSerAlaPr 788
Db      3254 ATTAATAATGCTACTTCTCGGGGTATATATCGCCGGATTTACGTGAATTTATATACAA 3313
OY      788 oCysIleProGlnGluThrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLys 808
Db      3314 AGAATTTGAAGAACTATCAATCACTAAAGAGTGCCTAATTTAGACTTAAT---GTATGCG 3370
OY      808 sAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeuProAsnGlyAlaTyr 828
Db      3371 TTATGATGCGTGAATGTCTGTGAATAATGATGAAAGCTTTGAAAATAACCGACTT 3430
OY      828 rIleLeuLysValAspGlyTyrThrThrLysIleAsn 840
Db      3431 C-----CAAGTAGATGTTTATGAAACAGATACTAAT 3461

```

RESULT 5
US-10-114-170-39
Sequence 39, Application US/10114170
Publication No. US20030023075A1
GENERAL INFORMATION:

APPLICANT: Blatner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Gay

WEICH, Rod
TITLE OF INVENTION: NO. US20030023075A1 Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESSES:
ADDRESS: Quarles & Brady
STREET: 1 South Plinkney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 8.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/114,170
 FILING DATE: 01-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/453,702
 FILING DATE: 03-DEC-1999
 APPLICATION NUMBER: 60/110,955
 FILING DATE: 04-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27386
 REFERENCE/DOCKET NUMBER: 960296,95017
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (608) 251-5000
 TELEFAX: (608) 251-9166
 INFORMATION FOR SEQ ID NO: 39:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 25165
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 39:
 US-10-114-170-39
 Alignment Scores:
 Pred. No.: 0.00135 Length: 25165
 Score: 145.00 Matches: 193
 Percent Similarity: 34.758 Conservative: 125
 Best Local Similarity: 21.098 Mismatches: 368
 Query Match: 3.27% Indels: 230
 Gaps: 48
 US-10-030-330-1 (1-843) x US-10-114-170-39 (1-25165)
 QY 67 GlyIlethrsErgIngluIngluYserProAlaTyrPheTyrValAlaAsnArg-GlyAs 86
 DB 2787 GCGGTAGCGGAATGCAAGCGGTGCTGCTGCAATGGAAAGACGCAACGCAATAC 2846
 QY 86 nAsngluYtYrAlaLeuValAlaAlaAspAspArgIleProThiIleuAlaTyrSe 106
 DB 2847 GGTCAACGGCGCGCAAGTACTAAGCGTCAATACCAACGCGCAACG---CTTGCCATCA 2903
 QY 106 rProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGlnI 126
 DB 2904 CACCATCGCTCAG-----GACACATT-----ATCAGTGC 2933
 QY 126 eTyrAspGlnIuIleGlyLeuIleuSerGly-----LysAlaGlnLeuAsnGluI 144
 DB 2934 TCGCGAACAATACGTCGCGCTGCTACTGAGCGGCAATGCGAATGCGAAGCGGGCAAC 2993
 QY 144 uIleLeuArgThrGluGlyValProAlaGluValAlaAlaLeuMetAspAsnGlyHisPh 164
 DB 2994 CGTACACTGACCGCTAACGGGAAAGCCATACAGCAACCGCTCGTAGCGAGCGA----- 3048
 QY 164 eAlaAsnAspPrometArgTrpAsnGlnGlyTyrProTrpAsnAsnLysGluProleu 184
 DB 3049 -----ACCTGGCAAGTACCGCTCGCTCCGACGGAAGTCCAG---GCACT 3089
 QY 184 uProAsnGlyAsnHisAlaTyrThrGlyCysValAla-----ThrA 198
 DB 3090 GCGCGAGGTAATATACGCTGTCATATGCGAGTGCACGCGATCGGCGAGGGAACCAACCA 3149
 QY 198 lAlaAlaGlnIleMetArg-----TyrHisSerTrpProLeuGlnGlyGluGlyS 215
 DB 3150 CCACACGGGAAATTCATTCAGTAGACACCTCAGACCCCGTGC----- 3190
 QY 215 ePheAspTyrHisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMet 235

DB 3191 -----TCACTGTATACCGTGGCGGACGAT----- 3220
 QY 235 yAspTrpIleAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValA 255
 DB 3221 -----TTCTAATAATATGCGAGCGCGCTCG-----CCGAGTCAATCTCCGG 3263
 QY 255 sPaLaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTyr---- 273
 DB 3264 ACAAGTCAGCGGTCTTCTCCAGCGCATACG---GTAACGGTAATATGGGCACTAGTCC 3322
 QY 273 ----- 273
 DB 3323 TGACGGCATCGCTGCGAGATGCGACGCTGGAATGTGGCGCTGGACCCGCGTAAACC 3382
 QY 274 -----GluAsnGlySerGlyThrTyrSerValTyrValGlyAlaLeuArgAsnA 291
 DB 3383 GCACGCTGATCGCGAGCGCAATACGATTTCTCACCCTGACAGATGCTGAGAAATA 3442
 QY 291 snPheArgTyrLysArgSerLeuGlnLeu---HisValArgAlaLeuTyrThr-----S 308
 DB 3443 CTGGCGCGCGCTCGACGATACGCTGCTGCTGCTTCTCCGTGATCACAATTAACA 3502
 QY 308 eArgGlnIuTrpHisAspMetIleArgGlyGluLeuAlaSerGlyArgProValTyrTyrA 328
 DB 3503 CCGCTCCGCGCGATGACATTATCAGTGC---GCAGAAAAGGTGGCGCCACTGACCTTA 3559
 QY 328 lAglyAsnAsnGlnSer-----IleGlyHisA 337
 DB 3560 CCGGTACACTCAACAGCGTGAAGAGCAAGCAACCGTCACAGTAACTCGCTGACAGA 3619
 QY 337 lAheValGlyAspGlyTyrAlaSerAspGlyThrPheHisPheAsn-----TrpG 354
 DB 3620 GTTTTACCACTACCGTGCAGGCG---GATGGCTCTGGAGTGCAGCGTACCTGCGCGC 3676
 QY 354 lYrPslGlyGlyLysSerAsnGlyPheTyrLysLeuThrIleu---LeuSerProThrSer 373
 DB 3677 CGATGGGAATATCGCTGACGCGCGCGTGGCGATTAACCGCTTCTGACGAGATCTACGC 3736
 QY 373 eGlyIleGlyGlyGluGlyIleGlyPheThrIle---TyrGlnIuIleIleThrGlyI 392
 DB 3737 GCAATACCGGCAACACTTCCCGCACCATTCACCTCGATACAGCGCGCGCTTAAGCA 3796
 QY 392 lGluProAlaLys-----ThrProAlaGluAlaGlyThrAspAlaLeuP 407
 DB 3797 TTGATCACTGACCGCTGATACATCATTAACCGCGCGAAGCGGCGAGAT---CTGC 3853
 QY 407 rIleLeuAlaLeuLysAspIleGlyAlaGluTyrLysSerGluSerGlyLeuAsn---- 425
 DB 3854 CCATACACCGGCAACCGACGCTACCGCGGCGGAGACGAGTACCGTTACCTTAATAGGC 3913
 QY 426 -----ValGlyTyrSerIleTyrA 432
 DB 3914 AGACGTATCAGGCGTCTGTCACGACAGCGCACTGAGACGCTGATGTCGCCGCCCA 3973
 QY 432 snThrGlyGluGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyG 452
 DB 3974 ACCTGGCGCA-----CTGGCTGACGCA 3997
 QY 452 lUValIleGluValLysThrSerSerIleAsnIleSerTrpTyrGlyTyrGlyLysHisP 472
 DB 3998 ACCT-----ACGGTACCGCCAGCGTGAAC-----GATGTCCCGGTAAATC 4039
 QY 472 rGluSerPheSer-----LeuAlaP 479
 DB 4040 CGAGCAGCGTTTACCGCTGCGCTGCTGATCCACGCCGCGGTGTAACCATTAATC 4099
 QY 479 rAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgArg-----ThG 497
 DB 4100 CCGTGGCGACCGATTAACGTATCAACACGCGGAGACATGCTCAGCGCAATATCAGCG 4159
 QY 497 lYThrGluGlnTrpGluProValArgHisAlaGlnGlyGlyTyrValAsnSerIleLysV 517

```

Db 4160 GCACG-----GTACTGCGCTCAGCGCGCATATCTACCGGTGACGC 4204
QY 517 aLaSnrThrThrAspProAsnAsnValValValThrValAlaPasnAsnGlnGlyLysLeuS 537
Db 4205 TGAATATATGTGAT-----TACACACACAGGTGTGATGTGATCGGCAACTGGA 4252
QY 537 erlle---ValProAsnSerPheValAlaAspleu-----AsnSerTyrglnHisSer- 553
Db 4253 GTCTGGCGCTCCGCGCTCGGTGTCAGTGGCTGGCGGACCGCATATCTCTGACACG 4312
QY 554 -----ThrlerThralGlnPheAsnS 561
Db 4313 TCTCGGTACCCAAACGCGGAAACACGCGGACCGCATCTTGACCGCTACCGTCAATA 4372
QY 561 eAspSerProAspGlnIleArgThrProValAlaPheAlaLeuSerThrlValAlaThra 581
Db 4373 CCGCGCGCGCC-----CTTACCGCATTAACGACATTCGCG---G 4408
QY 581 laAspAspValIleSerLeuGlyTrpValMetAlaGlnValPro---GlyGlySerSera 600
Db 4409 GCGATGATGTATTAAACCCGCAAGAAAGGCGCGATCTCCAGATTACCGGACCGACGC 4468
QY 600 snTyrrProValValTrpSerLysAspValLeuThrlLeuSerGlnGlyAspTyrrThrLeuT 620
Db 4469 ATCAGCCGTGTTACACCGCCATCAGCTG---ACCGTGAACGGCGCAAAATTTACACC--- 4521
QY 620 rPyrrArgPheSerIleAsnAsnGlnLysAspGlnIleLysLysIleGlySerValSerV 640
Db 4522 -----ACACAGACCGCGCGCTCCGCGCAACTG-----ACGC 4552
QY 640 allysThrProThrGlnGlyTrpThrlHisProLeuPheGlnValGlyHisAsnGlnThrSert 660
Db 4553 TCACCGCTCCGCGCATCGCGGTACAGATTAGGCGACGCCCACTATACGGTAACGCGCG 4612
QY 660 hrTyrrThrlLeuAspMet-----AlaHisAsnArgValLeuProAspPheT 675
Db 4613 CCGTGACACGAGCATATGCGCAACAGCGCCACTGCCAGCATACGCTGCTGCGACACGC 4672
QY 675 hrLeuLysAsnLeuGlyLeu---ProPheAsnGlnGlyLeuValValValAlaPheArgln- 693
Db 4673 CCGTCCCGCGGTGACATTAATCCGCGGCAACGAGCATATTTAACCCCGCGCG 4732
QY 694 -----ThrlnSerSerSerGlyLeuTrpAlaAlaGln-----GlnThrValH 709
Db 4733 CCGCGGTGCGCAACCATCAGCGCGGACGCTGCGCGGAGATGCGGACACGCGTAA 4792
QY 709 lsIleLysGlnGlyLysThrPheValTyrrLysProVal-----Y 722
Db 4793 CTATTACGTTGGGTGTAATCTTATACGCGACGCGGTGGACAAATCTCAGCTGACGC 4852
QY 722 aLgIuGlyProIleProAsp-----GlySerTyrrArgAlaThrLeuHisAlaP 738
Db 4853 TGGACGTTCACGCGGACATATTCAGCGCGCTGGGAAATGCGCATTTTAACTGTAATGCT 4912
QY 738 heValAsnGlnGlnGlnLeuTyrrLysGlnLysArgAsnTyrrThrlValLysIle- 757
Db 4913 CAGTCACCAATCAAAACGCGCAACGCGGACGCGGACGCGGATATCACCATCGACGCCA 4972
QY 758 --ValAsnGlyThrAlaValAlaGlnLysSerSerGlnGlnIleArgValPheProA 777
Db 4973 ATCTGCGCGCGCTGGCGGTGATACGCGGCGGCGGATGCTGCTCATATCATGACG 5032
QY 777 snProAlaArgAspTyrrValGlnIleSerAlaProCysIleProGlnIleThrSerIleT 797
Db 5033 ACGGCGACGCGCTGTGTGTACCGCGACAGCTGCGGCTGTGAAAGCACGCGCGCTTA 5092
QY 797 leLeu-----PheAspLeuSerGlyLysIleValMetLysAsnSerLeuSera 813
Db 5093 CCGTTACGATTAAATATGTGGAATACACACTGCGGTGACGCGCATGATGATGATGAGCG 5152
QY 813 laGlyHisGlyArgMetAspValSerArgLeuProAsnGlnAlaTyrrIleLeuLysValA 833
Db 5153 TGGCGGTACGCGCGCGCGAGGTATGCGCGCTGCGCGGACGTTAAATATGCGCGTTT 5212

```

```

QY 833 spGlyTyrrThrThr-----LysIleAsnIleValHis 843
Db 5213 CAGCGGAAAGTACGCGCGGAAACTCGGTGACGATTAACGAT 5253

RESULT 6
US-08-837-459-21
Sequence 21, Application US/08837459
Patent No. US20020006407A1
GENERAL INFORMATION:
APPLICANT: McKee, Marian L.
APPLICANT: O'Brien, Allison D.
APPLICANT: Machtel, Marian R.
TITLE OF INVENTION: Histidine-Tagged Intimin And Methods Of
TITLE OF INVENTION: Using Intimin To Stimulate An Immune Response And As An
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flanagan, Henderson, Parabow, Garrett &
ADDRESSEE: Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/837,459
FILING DATE: 18-APR-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04995, 0023-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 3106 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-837-459-21

Alignment Scores:
Pred. No.: 0.000188 Length: 3106
Score: 140.00 Matches: 167
Percent Similarity: 30.67% Conservative: 128
Best Local Similarity: 17.36% Mismatches: 331
Query Match: 3.15% Indels: 336
Gaps: 45

US-10-030-330-1 (1-843) x US-08-837-459-21 (1-3106)
QY 74 GlySerProAlaTyrrPheTyrrValAlaAsnArg-----GlyAsnAsnGlnGlyTyrrAla 91
Db 144 GGTTCGATGCTGTTTATGCTTATATCATGAAATCATTTGCAAAATGCGAAATTAATTTT 203
QY 92 LeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyrrSerProIleGlyArgPhe 111
Db 204 AAATGGGTTCGATTAACAACTGTAACTATCATATGATATGATCAG-----AATCGCCTT 257
QY 112 -----AspMetAspSerMetProAsp-----Asn 119
Db 258 TTTTATACGTTGAAAACTGTGAACACTGTGCGCATCTTTCTAAATCGCAAGATTAATTAAT 317
QY 120 Leu---ArgMetTrp-----LeuGlnIleTyrrAspGlnIleGlyLeuIleLeu 135

```

```

Db      318  TATGACGATTTGGCTGTAATAGCATTTATACGTTTCAAGGCAAGATGATGAG 377
Oy      136  SerGlyValAlaThrLeuAsnGluGluLeuAlaGlyThrGluGlyValProAlaGluVal 155
Db      378  GCGGCGCCCTGGTCAG-----CAGATCATTTTCCCACTCAAAAAAATCCCTTGAATAC 431
Oy      156  HisAlaLeuMetAspAsnGlyHisPheAlaAsnAspPheMetArgTrpAsnGlnGlyTrp 175
Db      432  AGTCAGCTA----- 440
Oy      176  ProTrpAsnAsnGlyGluProLeuLeuProAsnGly----- 187
Db      441  CCACATTTTAGTTCGGCACCTCTGTGTGTCAGAGGCGTTCGTGTCACACGATTA 500
Oy      188  ----- 501
Db      501  CTGACTAAATATGTCCTCCGACGTGACCAAAACACATGACCGATGACAAAGCATTA 560
Oy      199  GlyCysValAlaThrAlaAlaGlnIleMetArgTrpHisSerTrpProLeuGlnGly 212
Db      561  TATGCGGCACAAACAGCGCGAGTCTCGGTAGCCAGCTTCAGTCGCGATCTCGAAGGC 620
Oy      213  GluGlySerPheAspTrpHisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGly 232
Db      621  GATTAACGCAAGATATACGCTCTGTGT-----ATCGCTGTAACACAGCTTCGTCA 671
Oy      233  GluMetTrpAspTrpIleAsnMetProGlyAsnProAspLeu-----Asp 247
Db      672  CAGTTGACAGCGCTGTACACATTATGACAGCGCATTAATCGACAGATGGTGAT 731
Oy      248  AsnLeuThrGlnSerGlnValAspAlaTrpAlaThrLeuMetArgAspValSerAlaSer 267
Db      732  AACTTTGACGGTAGTCTACTGAC----- 755
Oy      268  ValSerMetSerPheTrpGluAsn----- 275
Db      756  TTCCTATTACGCTTCATGATTCGCAAAAAAATGCTGGCATTTGTCAGGTCGAGCGCT 815
Oy      276  -----GlySerGlyThr----- 279
Db      816  TACATTCACCTCCGCTTACGGCAAAATTTAGTGGCGGTACAGCTTTTCTCTCTGCA 875
Oy      280  -----TyrSerValTyrVal----- 284
Db      876  AACATGTTGGCTATTAACGCTTCATGATCAGGATTTTCTGGTGATTAATACCGCTTA 935
Oy      285  ---ValGlyAlaLeuArgAsnAsnAsnPheArgTrpLysArgSerLeuGlnLeuHisValArg 303
Db      936  GGTATTGGTGGCGAATACGGCGAGACATTTTCAAAAGTAGC-----GTTAAC 983
Oy      304  AlaLeuTrpHisSerGlnGluTrpHisAspMetIleArgGlyGluLeuAlaSerGlyArg 323
Db      984  GGCCTATTTCCGATGAGCGCTGGCATGAGTCATACCATTAAGAAACACTATGATGAGCC 1043
Oy      324  ProValTyr-----TyrAlaGlyAsnAsnGlnIleSerIleGly 335
Db      1044  CCAAGCAATGCGCTTGATATCCGTTTAAATGGCTATCTACCGCTCATATCCGGATTAAGC 1103
Oy      336  HisAlaPheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTrp 355
Db      1104  GCCAAGCTGATATATAGACGATATATGATATGATTAATGCTTGTGTTAATCTGAT 1160
Oy      356  GlyGlyValSerAsnGlyPheTrpLysLeuThrLeuLeuSerProHisSerLeuGlyIle 375
Db      1161  -----AAGCTG-----CAGTCGAATCCTGCTGCG 1184
Oy      376  GlyGlyGluGlyIleGlyPheThrIleTrpGlnGluIleIlePheGlyIleGluProAla 395
Db      1185  GCGACCGCTGGTGTAACCTATACCTCGATTCCTCTGGGTACAGATGGGATC----- 1235
Oy      396  LysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlu 415

Db      1235  ----- 1235
Oy      416  AlaGluTrpLysSerGluSerGlyLeuAsnValGlyTyrSerIleTrpAsnThrGlyGlu 435
Db      1236  ---GATTACCGT-----CATGTACGGGTAAAT 1259
Oy      436  GluGlnSer-----AsnLeuAspLeuGlyTyrArgLeuAsnGlyValAspGlyGlu 452
Db      1260  GAAATGATCTCCTTTACTCATGACGTTCCGTTATCATGATTAATGATGCGTCTCAG 1319
Oy      453  ValIleGluValLysThrSerSerIleAsnIleSerTrpGlyGlyGlyGlnIlePro 472
Db      1320  CAATATGAAACACAG-----TATGTAAACAGGTA 1349
Oy      473  GluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeu 492
Db      1350  AGAACATTTATCAGGACCGGTTACGATCTGTTCACCGTAATATATATTCGCGAG 1409
Oy      493  TyrArgArg-----ThrGlyThrGln 500
Db      1410  TACAGAACAGAGATATCTTCTGTGAATATTCGCCATGATTAATGATGTAACAC 1469
Oy      501  -----Trp 501
Db      1470  AGTACGCAAGATTCAGTTGATCTTAAGACCAATACGGTCTGATGTCGTCTGG 1529
Oy      502  ---GluProValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsnThr 520
Db      1530  GATGATAGTCATTTACGACGAGTCAGGCGGTACAGATTACACATGACCGCAACACGCA 1589
Oy      521  AspProAsnAsnValAlaThrValAspAsnAsnGlyLysLeuSerIleValPro 540
Db      1590  CAAGACTACACAGCTATTTTGCTCTGTATGTCGAAGTGGCAGC----- 1634
Oy      541  AsnSerPheValAlaAspLeuAsnSerTrpGluHisSer----- 553
Db      1635  AATATTATTAAGTAGACGGCTCGGCTTATGACCGTAATAGCGTAACATGTA 1694
Oy      554  ---ThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThrProVal 571
Db      1695  CAGCTTACATTAACCTCTGTGCAATGTCAGATGTGTCAGCAGGCTGGGTAAAGCGAC 1754
Oy      572  AlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTrpValMet 591
Db      1755  TTTACGCGGATTAACCTTCGGCTAAAGGGAATACCGGATACATTAATCTTATACCGG 1814
Oy      592  AlaGluValProGlyGly-----SerAsnTrpProValTyrPheLysAspVal 609
Db      1815  ACGGTGAAAAGAAAGGGGTAGCTAGCTGATATGCTCCGTTCATTTAAT 1865
Oy      610  LeuThrLeuSerGluGlyAspTrpThrLeuTrpTrpArgPheSerIleAsnAsnGlnLys 629
Db      1866  ---ATTGTTACAGAACTGCACACTT-----GGGCAAAATATGTCAAA 1907
Oy      630  AspGluTrpLysLysIleGlySerValSerValLysThrProGlnGluTrpHisPro 649
Db      1908  ACGGATGCTAACGCTTAAGCAACGTTACGTTAAATCCAGTACGCCA----- 1955
Oy      650  LeuPheGluValGlyHisAsnGlnThrSerThrTrpThrLeuAspMetAlaHisAsnArg 669
Db      1956  ---GACAGGTGCGTGTCTGCTGCTAAACCGGAGATGAGTTCA----- 1997
Oy      670  ValLeuProAspPheThrLeuLysAsnLeuGlyLeuProPheAsnGlyGluLeuVal 689
Db      1998  -----GCACTTAATGCCAGTCGCGTTATA 2021
Oy      690  ValPheArgGlnThrGlnSerSerGlySerLeuTrpAlaAlaGlnThrValHis 709
Db      2022  TTTTGTGATCAACCAAGCGCACGATTAATGATTAATGATTAATGATTAATGATTAAT 2078
Oy      710  IleLysGlnGlyGlu-----ThrPheValTyrLysProValValGluGly---Pro 725
Db      2079  GTACGAAATGTAAGATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2138
```

```

OY 726 IleProaspGlySerTyrArg---AlaThrLeuHisAlaPheValAsnGlyGlnGln 744
DB 2139 GTTATATATCATCCGTTATCATCTTCACAACTTGGATGTCACAGGTAAGTCTCA 2198
OY 745 LeuTyrLeu-----Lys 748
DB 2199 ACGCAAGCAACCCGGAATGATGTCGCGACGATPACACTCACTTCAGTTCGCCG 2258
OY 749 GlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaValGluAlaIleGluSer 768
DB 2259 GGTAAAGCAGCTGTTAGTCCGACAGTCACTGATGGGCTGAGGTTAAAGCGAGTGC 2318
OY 769 Ser-----GluGluIleArgVal----- 774
DB 2319 ACTTTTGTGATGAACTGAAATGTGACAAAGTGTGATATTGTTACATGTCACA 2378
OY 775 -----PheProAsnProAlaArgAspTyrValGluIleSerAlaProCys----- 789
DB 2379 GCGGAGTTCCTTAATATTGGCTGCATATGCTTAACTGAAGCAAGCGGTGCT 2438
OY 790 -----IleProGlnGluThrSerIleIleLeuPheAspLeuSerGly 803
DB 2439 GATGCTACATATTCATGCTATTCAGAAATACAGATGCGACGTGTCATCATCAGG 2498
OY 804 LysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeu 823
DB 2499 AAGGTCATCTTGAAT-----GGTAAAGCGAGTCTC---GTAATTAAAGCC 2540
OY 824 ProAsnGly-----AlaTyrIleLeuLysValAspGlyTyrThrThrLys 838
DB 2541 ACATGTGGATGATACCAACAGTAACTACTATMAAGCACCGCTGTTATGATMAA 2600
OY 839 IleAsn 840
DB 2601 GTGGAT 2606

```

RESULT 7
 US-10-114-170-18
 ; Sequence 18, Application US/10114170
 ; Publication No. US20030023075A1
 GENERAL INFORMATION:
 APPLICANT: Blattner, Frederick R.
 Burland, Valerie
 Perna, Nicole T.
 Plunkett, Guy
 Welch, Rod
 TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli 0157
 NUMBER OF SEQUENCES: 265
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Quarles & Brady
 STREET: 1 South Plinckney Street
 CITY: Madison
 STATE: WI
 COUNTRY: US
 ZIP: 53701-2113
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 8.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/114,170
 FILING DATE: 01-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/453,702
 FILING DATE: 03-DEC-1999
 APPLICATION NUMBER: 60/110,955
 FILING DATE: 04-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27386

```

REFERENCE/DOCKET NUMBER: 960296, 95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3695
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-114-170-18

Alignment Scores:
Pred. No.: 0.000277 Length: 3695
Score: 139.50 Matches: 211
Percent Similarity: 31.32% Conservative: 111
Best Local Similarity: 20.53% Mismatches: 372
Query Match: 3.14% Indels: 336
DB: 9 Gaps: 47

US-10-030-330-1 (1-843) x US-10-114-170-18 (1-3695).
OY 63 AspaIaGluArgGlyIleThrSerGlnGlnGlySerProAlaTyrPheTyrValAla 82
DB 526 GACGGTCAGAGCCGATTAATCAATGCTGCTAAGCGTAATGAGCAAAACCGGTGCTG 585
OY 83 AsnaTgGlyAsnaGlnGlyTyrAlaLeuValAlaAspAspArgIleProThrIle 102
DB 586 TCTGCGCGAGCCGCGGAGGCGACGACGTACAGGCGCATGAAATCATAGATCAAGCTAA 645
OY 103 LeuAlaTyrSerProIleGlyArgPhe-----AspMetAspSerMetProAspAsn 119
DB 646 CTAACCTTCAAAACCGCGTGAATATGTGACCTGCTCCCTGAAGCCACATAATCACAG 705
OY 120 LeuArgMetTyrPleu---GlnIleTyrAspGlnGluIleGlyLeu----- 133
DB 706 GCAAAGCCAAACACTGGGTGAGTTCCACCGAACTGAAGCAGGCGGTATCATGCTGCTTT 765
OY 134 -----IleLeuSerGlyLysAlaGlnLeuAsnGlnGluIle---LeuArgThrGlu 149
DB 766 ACTACCGGAACGCTAGTGAAGGCAACGATTAAGCTTACGCTGATGACGATGAGCAAA 825
OY 150 GlyValProAlaGluValHisAlaLeuMetAspAsnGlnHisPheAlaAsnAspPromet 169
DB 826 ACCGTCACTGCGAAGTACGCGGCCACAGTATGAT-----GTGCAAACTCCACCCCTG 879
OY 170 ArgTyrAsnGlnGlyTyrProTyrAsnAsnLysGluPro-----LeuLeuProAsn 186
DB 880 AGCGCTAAC-----GAGCCGTGACGTGACGTGCTGCTGAT 915
OY 187 GlyAsnHisAlaTyrThr-----GlyCysValAlaIle 197
DB 916 GGTACAGAACCTATACGTTGACGTGACGTGCGGTGACCTCCGAGGTAATCCGGTGACG 975
OY 198 AlaAlaAlaGlnIleMetArgTyrHisSer-----Tyr----- 208
DB 976 GGAGAAGCAGCCGCTTGCGATTTGT-TCGGAAGACACTAATAGTGTAAACCGTTGGTGC 1034
OY 209 -----ProLeuGlnGlnGlySerPheAsp----- 217
DB 1035 CATTCGGAATTAACACAGCGGTTTACAGCGCCCGGTTCTTGACCCGCGCGA 1094
OY 218 CGTTGTGTGCGTCTTACAGGACAGTATCAGCTGGGCAATTACAAACAACCGCTGAA 1154
DB 1095 GGTGTGTGCGTCTTACAGGACAGTATCAGCTGGGCAATTACAAACAACCGCTGAA 1154
OY 224 -----Val 224
DB 1155 GTTGTGCGCGCGCTTATGACGACACATTCGTCATCACCTGGAATCCGTATAAAC 1214
OY 225 GlyAsnTyrSerGlyThrPheGlyLysMetTyrAspTyrIleAsnMetProGlyAsnPro 244

```

```

Db 1215 GGTGGTGGGGAGAGTACGGC-----ANCTGGACGGTAAGA 1256
Qy 245 Aspleu-AspAn-----LeuthrInserInValAspAlaTyrAlaThrLeuMetArgAs 263
Db 1257 TGCCATATGACCACTGTGACAGCCCTCAACGCCCGCGCTCATTTAGCGGTGCCG 1316
Qy 263 pValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyrSerValTy 283
Db 1317 TGCTAAGAGTTTACGGCATCGGGCTGGCAATAATATGGTATGGAGCTGGACGTGGCA 1376
Qy 283 rValValAlaLeuArgAsn-----AsnPheArgTyrLysArgSer----- 297
Db 1377 GATTACTCTCGGCTCTACGGCGGGTGAATTAGAGTTATGCCAAGCTTAATGACAGAA 1436
Qy 298 -----LeuGluLeuHis-----ValArgAlaLeuTyrThrSerG 309
Db 1437 TGCGGCACCAATCGCGGCAAAAGTAACCTGGTGGCTGATGGCTATCTTCAAAACCAATC 1496
Qy 309 InGU-----T 311
Db 1497 GAAATCTCTGTGCGAAGATACAGTAATAACCGCGCAACACACACCGTACGCTGGT 1556
Qy 311 rPhIsAspMetIleArgGlyLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnA 331
Db 1557 GCGCAAAAGATGCG-CATGGCAACGCTATCAGTGT----- 1590
Qy 331 sngInSerIleGlyHisAlaPheValCysAspGlyTyrAlaSerAspGlyThrPheHisP 351
Db 1591 -----CTTGGCTGTGGCGCAAGTTTGACCGGGGCGCTCTGAAAGGGGCGCGTTT 1642
Qy 351 heAsnTrpGlyTyrPglYgLYValSerAsnGlyPheTyrLysLeuThrLeuSerProT 371
Db 1643 CCAGTTGACCGCAAAAGGT-----AACGTTCTATGTGCTACGTTGACTACAGGTG 1696
Qy 371 hrSerLeuGly----- 374
Db 1697 GAAAGACGGGCGAGCTTGCCTCATGCTCTCTTCACAGGCCAGCAGCAGCAGCAGG 1756
Qy 375 -----IleGlyGlyGlu----- 378
Db 1757 CCGCGACAGTTGACGGTATCGCGGAGAGATGTCATACAGCAACTCTACGCTTGTTCGG 1816
Qy 379 -----GlyIleGlyPheThrIle-----T 385
Db 1817 ACAATTAAGCTCCGACCGTCAAACAGCAGCAGCAACTCCTTACCGTGAAGAGATGCGT 1876
Qy 385 yrgInGUleIleIleThrGlyIleGluPro-----AlaLysT 397
Db 1877 ACGGGAACCCGGGTACCGGGGTGAAGCCAGATGCACACAGTGTTAGCGGTGGCGCGACA 1936
Qy 397 hrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluAlaG 417
Db 1937 CGGGAGTGAGGCTCTTCA-----GCAG 1960
Qy 417 LuTyrLysSerGluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGluG 437
Db 1961 GAAATGACAGCAAGAAAGTAATGGGCTTACGTTGACCTTAACCGTGGAGTCTGCCG 2020
Qy 437 InSerAsnLeuAspLeuGlyTyrArgLeuAsn----- 447
Db 2021 CGGGTACGTTGCTGTGATGCGCGAGTGAAACGGCAAAATGCCGTTGCTCAGCCATGG 2080
Qy 448 -----LysAlaAspGlyGluValIleGluValLysThrSerIleAsnLIS 464
Db 2081 TGCTGAACGTTGACAGTGCAGCATTAAGGCTGAGATTCGATATGACAGTGAAGGTTA 2140
Qy 464 eTrTPtyrGlyTyrGlyGluHisProGluSerPheSerLeuAlaProAsnGluSerG 484
Db 2141 AT-----AACCACTGGCTA 2155
Qy 484 InGU-----IleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGluInTrpG 502
Db 1215 GGTGGTGGGGAGAGTACGGC-----ANCTGGACGGTAAGA 1256
Qy 502 LuProValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsnThrAsp 522
Db 2185 --ACCGTTGACACCTATGTAAACCCGTGGAGGGCAGAGATAGCTGTTTAC 2242
Qy 522 roAsnAsnValVal-----ValThrValAspAsnAsnGluGlyLysLeus 537
Db 2243 CGCAGGCTGTGACACCAAGCAGCGGGAATACATAACCACTAATTCGGCGCAATMAACGG 2302
Qy 537 eTrLeValProAsnSerPheValAlaAspLeuAsnSerTyrGluHisSer----- 553
Db 2303 ACATTAAGCTTATGTCACAGGTTGCCGGAACACACATATTTCCGCTTCCGTAATGCTG 2362
Qy 554 -----ThrIleThrValGlnPheAsnSerPhe-----SerProAspGluIleArgThr 570
Db 2363 CTCAGAAAGACGGCAGCGGTAATTCACCGGATCCACACCGGTCAGCAGCAACCTGC 2422
Qy 570 roValAlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTyrV 590
Db 2423 AGGTACAGCCCGCTGCTCAAAAGTGCAAAACGGCAAAAGATGCTTACCTGACGGCA 2482
Qy 590 aLMetAlaGlu-----ValProGlyGly-----SerSerAsnTyrPro- 602
Db 2483 ACCTTGAGTAAATAATGTAACTGTTCCAGGAGACCTGTGACCTTAATCTGCCCC 2542
Qy 603 -----ValValTrpSerLys----- 607
Db 2543 GGGGTGTCAAGCCGCTTACAGCGGATATCTGTGGTGAAAGCCAAAGATGAGGGGAAG 2602
Qy 608 -----AspValLeuThrLeuSerGlyGlyAspTyrThrIleTyrArgPheSerI 625
Db 2603 CAGAGTTGACAGGCGTTTACAGTACGTCGGCAAGCTATGATCAGCGCATCGGCAGGG- 2661
Qy 625 leAsnAsnGluLysAspGluTyrPlyLysIleGlySerValSerValLysThrProThrG- 645
Db 2662 --AATAGCCAGCCTTGAAATACAGCAGCTTAAAGCTTTGAGCGGAATGCTACCGCA 2719
Qy 645 LuTyrThr-----HisProLeuPheGluValGlyHisAsnGlnT 658
Db 2720 CCGTCTCCGCTATTTAGGTGATGGAACACTATCTGCGGCGCAATGCGCAACAGCA 2779
Qy 658 hrSerThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAsp-----PheThrL 676
Db 2780 CGTATAAGTTACGGTGATGATGCCAATTAACCACTGTTGAAGATAGCGAAGTGAAGC 2839
Qy 676 eu-----LysAsnLeuGlyLeuProPheAsnGlyLeuValValPhea 692
Db 2840 TGACTGCCAGCCCGCAAAATTTAGTTCTGACTCCCAATGGGAGC-----G 2884
Qy 692 rGlnThrGlnSerSerSerGlySerLeuThrAlaAlaGlnGluThrValHisIleLysG 712
Db 2885 CGAAACCTAATGACGAAGCAGGCTAATTTTCCGCCGACGACACTGTGTCGACGCAAA- 2943
Qy 712 InGUleuThrPheValTyrGlyProValValGluGlyProIleProAspGly-----S 730
Db 2944 -----TATACSTCAACCGCGCAAAAGTGAAGTACAGCGCGGCGAGAGAT 2986
Qy 730 eTrTyrAlaThrIleuHisAlaPheValAsnGlyGlnGlnGluLeuTyrLeuLysGlyL 750
Db 2987 CGACGAAACSTGCGCAATCTAATTTGCTGCGGAGATACAAATGCAAGTACAGCCGAT 3046
Qy 750 yAsArgAsnTyrThrValLysIleValAsnGlyThrAlaValAlaIleGluSerSerG 770
Db 3047 CATCTGATGTGACTTCTGTGTCGCGAGTGGATATCAGCTGCGAAGCTGAGAGTACAC 3106
Qy 770 InGUleuLeuValPheProAsnProAlaArgAspTyrValGluIleSerAlaPro----- 788
Db 3107 TGATGTGCGCAATTAACCCGTTGGGGGGAATATGTGGGTGCAATTAAGCCCGCAAG 3166
Qy 789 -----CysIleProGlnGluThrSerIleIleLeuPheAspL 801
Db 3167 GGTGACGGAAGATTAATCACTGCTCCGCGCAAAATGACCAATTC-----G 3217
```


QY 801 euSerGlyValIleValMetLysAsnSerLeuSer---AlaGlyHisGlyArgMetAspV 820
DB 3218 TACGGCGAATAATACAGCGCTACATTTAGTACAGCAGCGCTGTCTACTGTTACAT 3277
QY 820 aSerArgLeuProAsnGlyAlaTyrIleLeuLys-----V 832
DB 3278 TTAAGCCCTGACGTATGCGGTACGAATAATGAGCAGTACAGTACAGTATACCGCG 3337
QY 832 aLAspGlyTyrThrThrLys 838
DB 3338 TGAATGCGATACGCGCAAG 3357

RESULT 8
US-10-080-505-10
Sequence 10, Application US/10080505
Publication No. US20030073166A1
GENERAL INFORMATION:
APPLICANT: St. Geme, Joseph W.
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
REFERENCE: A-59941-1/RET/DCF/DHR
CURRENT APPLICATION NUMBER: US/10/080,505
CURRENT FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: US 08/296,791
PRIOR FILING DATE: 1994-10-25
PRIOR APPLICATION NUMBER: US 09/839,996
PRIOR FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 4605
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: CDS
LOCATION: (422)..(4597)
OTHER INFORMATION:
US-10-080-505-10

Alignment Scores:
Pred. No.: 0 000494 Length: 4605
Score: 138.50 Matches: 165
Percent Similarity: 31.36% Conservative: 89
Best Local Similarity: 20.37% Mismatches: 275
Query Match: 3.12% Indels: 281
DB: 9 Gaps: 44

US-10-030-330-1 (1-843) x US-10-080-505-10 (1-4605)

67 GlyIleThrSerGlnGluGlySerProAlaTyrPhe-----TyrVal 81
DB 473 GGGATAGTATCGCAAGCGTGGGAGGTCTACTTATTGGGATTGACTACCAATTTAT 532
QY 82 AlaAsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAsp----- 97
DB 533 CGTGTATTTGCGGAGATAAGGAAGTTTACAGTTGGCGCTCAAGATTTGATATCTAC 592
QY 98 -----ArgIleProThrIleLeuAlaTyrSerProIleGlyArgPheAsp 112
DB 593 AATAAAAAGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 646
QY 113 MetAspSerMet----- 116
DB 647 TTAATCTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 706
QY 117 -----ProAsp 118
DB 707 GTCCGACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 766
QY 119 AsnLeuArgMetIlePheGlnIle-----TyrAspGlnGluIleGlyLeuIleLeu 135
DB 767 CAACATCGTTTAACTTAAAGTTGTTAAAGCATATTAATTATAG----- 811

QY 136 SerGlyLysAlaGlnLeuAsnGlu-----GluIleLeuArgThrGlu 149
DB 812 AGCGGTGATAGCAT 871
QY 150 GlyValProAlaGluValHisAlaLeuMetAspAsnGlyHisPheAlaAsnAsp----- 167
DB 872 ACTGCACCTATTTGAATGCTTTCAATATATGATGATGATGATGATGATGATGATGATGATGAT 931
QY 168 -----ProMetArgTyrAsnGlnGlyTyrProTyr-----AsnAsnLys 180
DB 932 TATCTTTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 991
QY 992 -----ACCATGACACTTACGCTTATGAGCT----- 1018
DB 201 GlnIleMetArgTyrHisSerThrProLeuGlnGlyGlySerPheAspTyrHisAla 220
QY 1019 -----TCATGTTTAAATAGGTGGAATACCTTTGAAGATGACACCGCT 1060
QY 221 GlySerLeuValGlyAsnThrPheGlyThrPheGlyGluMetTyrAspThrLeuAsnMet 240
DB 1061 GGTATC----- 1066
QY 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrThrLeu 260
DB 1067 -----GGTACATATGATTAATATGCGGAGTACCAAAATCC-----ATAAATATGTCACCA 1120
QY 261 MetArgAspValSerAlaSerValSerMetSerPheTyrGlnAsnGlySerGlyThrTyr 280
DB 1121 -----CCTACGCGACGCTTATTCGGGATGAGTGTCTCCAAATG--- 1159
QY 281 SerValTyrValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeu 300
DB 1160 -----TTATTTATGATTAAGCACTT----- 1180
QY 301 HisValArgAlaLeuTyrThrSerGlnGlyTyrHisAspMetIleArgGlyLeuAla 320
DB 1181 -----AAGAAATG-----TTATTAATGCGCTTACGT 1210
QY 321 SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 340
DB 1211 GAAGGAATTCCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1270
QY 341 AspGlyTyrAlaSer-----AspGlyThrPheHisPheAsnTyrGlyTyrPheValSer 359
DB 1271 CAAGGTATTTCAATCAAGACATTTACAGCTAATTT-----TGGAT-----ACT 1315
QY 360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGly 379
DB 1316 AATGCTGATATATGATTT-----AATATGAGGAGTACCAATGGA 1357
QY 380 IleGlyPheThrIleTyrGlnGluIle-----IleThrGlyIleGluProAlaLysThrPro 398
DB 1358 AGAGTGGCAACATCAAAAGTACATTAACCTTAAAGCAATTAATTAATTAATTAATTAATTAAT 1417
QY 399 AlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLys-----AspIleGluAla 416
DB 1418 GGTCTTTATGATTAATAGCACTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1477
QY 417 GluTyrLysSerGluSer-----GlyLeuAsnValGlyTyrSerIleTyrAsn 432
DB 1478 TCTTATTAAGGTCATATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1537
QY 433 ThrGlyGluGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlu 452
DB 1538 GCGGATCAAGGAACAGCACTTAACATTTGAATAATTAATTAATTAATTAATTAATTAATTAAT 1597
QY 453 Val-----IleGluValLysThrSerSerIleAsnIleSerTyrGly 467
DB 1598 TTGTATTTTGAAGTAAATTTTGTGTAAGGCAATCAAAATTAATTAATTAATTAATTAATTAATTAAT 1657
QY 468 TyrGly-----GluHisPheGluSer 474

```

Db      1658 GCAGGCGTTCTGTTGGAGAGAAAGTACTGTTGATGCGAGGTGCATATATCCAGAGGC 1717
      475 PheSerLeuAlaPro-----
Db      1718 GATCGCTTATCCAAATTTGGCGTGGACCTTACTGTTAATGCTAAAGGAAACTTA 1777
      480 AsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGlu 499
      1778 GGAAGCGTAGTGGTGGTACGGTTTGGTGTGA-----GATCAACAGCAGATGAA 1831
      500 GlnTPGluProValArgHisAlaGlnGlyIleTyrValAsn-----SerIleLys 516
      1832 TCAGGTCAAAAACAGACCTTTAAAGAGTGGCATGTGTAAGTGTGAGCTACCGTCA 1891
      517 ValAsnThr-----ThrAspProAsnAsnValVal-----Thr 528
      1892 CTAATATAGTCAGATCAAGTGTGATCTCAATATTATTTGCGCTTTCGTGGTGGCC 1951
      529 ValAspAsnAsnGlnGlyLysLeuSerIle-----ValProAsn-----Ser 542
      1952 TTAGATCTTAATGGGCATTCATTAACCTTTGAACGATTCCAAATATCGATGAAGCGCG 2011
      543 PheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrVal-----Gln 558
      2012 ATGATTTGTCACCAACAGCTTCTCAACCGCAATATATACGATTACAGCAACGCACT 2071
      559 PheAsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGly 578
      2072 ATTAATTCAGATGAGCAACACTT-----
      579 AlaThrAlaAspValIleSerLeuGlyTyrVal-----MetAlaGluVal 594
      2096 ACTAATTAATAAAGATTTTCATTTAAGCGCTGTTGGTGAGCAGATTAAGCTTAACA 2155
      595 ProGlyGly---SerSerAsnTyrProValIleTyrSerLysAspValLeuThrLeuSer 613
      2156 AATGCTGCTTAATGTAATGTAATATCAACAGATTAAAGCAAAATCATTTGTTGCTTCT 2215
      614 GluGlyAspTyrThrLeuTyrPyrArgPheSerIleAsnAsnGlnLysAspGluTyrLys 633
      2216 GGGGG-----ACAAATTTAAACGCAATATACAG 2245
      634 LysIleGlySerValSerVal-----LysThrProThrGluTyrThrHisPro 649
      2246 CAAATGCGTGAACCTTAGTTTTAGTGTGCTCCAGCCCTCATGCTTACATCAT--- 2302
      650 LeuPheGluValGlyHisAsnGlnThrSerThrTyrThrLeuAspMetAlaHisAsnArg 669
      2303 -----TTAAGAAGAGACTTGTCTAATGAGAA 2329
      670 ValLeuPro-----AspPheThrLeuLysAsn 678
      2330 GGTATCCCAAGGCAATTTGTGGGATCAGATGTGATCAACCGCATTTAAAGCT 2389
      679 LeuGlyLeuProPheAsnGlnGlyLeuValValValPheArgGlnThrGlnSerSerSer 698
      2390 GAAACTTCCAAATTAAGCGGGAAGTGGGTGTTCTCGCAATGTTTCTTCAATTAAG 2449
      699 GlySerLeuTyrAlaAlaGlnGluThrValHisIleLysGlnGlyLysThrPheValTyr 718
      2450 GGAAT---TGG-----ACAGCAGCAATATGCAAAATCCCATTT----- 2488
      719 LysProValValGlnGlyProIleProAsp 728
      2489 -----GGTGTGTGCCAAAT 2503

```

RESULT 9
 US-10-080-505-14
 ; Sequence 14, Application US/10080505
 ; Publication No. US20030073166A1
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Geme, Joseph W.

```

; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-509641-1/RET/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080, 505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 4822
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (388)..(4563)
; OTHER INFORMATION:
US-10-080-505-14

Alignment Scores:
Score: 0.00053 Length: 4822
Percent Similarity: 138.50 Matches: 165
Best Local Similarity: 31.368 Conservative: 89
Query Match: 20.378 Mismatches: 275
DB: 3.128 Indels: 281
Gaps: 44

US-10-030-330-1 (1-843) x US-10-080-505-14 (1-4822)
QY 67 GYIleThSerGlnGluGlySerProAlaTyrPhe-----TyrVal 81
Db 439 GGGATGTATCGACACCGGCGCAGCTACACTTATTTGGGATTGACTACCAATTTAT 498
QY 82 AlaAsnArgLysAsnGlnGlyTyrAlaLeuValAlaAlaAsp--- 97
Db 499 CGTGAATTTGCCAGATTAAGGGAAGTTTACAGTTGGGCGCAAGATATGATCTAC 558
QY 98 -----ArgIleProThrIleLeuAlaTyrSerProIleGlyArgPheAsp 112
Db 559 AATTAATAAAGGGAATGATAGTACGATGATGAAGTGTGCTATGCT---GAT 612
QY 113 MetAspSerMet----- 116
Db 613 TATCTTCATGATGTTGGTGTGCTATTCACATGTATAGTACGACATTAATTAAC 672
QY 117 -----ProAsp 118
Db 673 GTCCACATTAATGTAGGTATGATGTCGTTGATTTGGTATGGAGGGGAAATCCAGAC 732
QY 119 AsnLeuArgMetThrPheGlnIle-----TyrAspGlnGluIleGlyLeuIleLeu 135
Db 733 CAACATCGTTTAAAGTATAAAGTTGTTAAACGATATATATTAAG--- 777
QY 136 SerGlyLysAlaGlnLeuAsnGlu-----GluIleLeuArgThrGlu 149
Db 778 AGCGGTATGACCAATATATATGTTATCAACATCCCAAGATTAGAGAAATTTGTAACGAA 837
QY 150 GlyValProAlaGluValHisAlaLeuMetAspAsnGlnHisPheAlaAsnAsp----- 167
Db 838 ACTGACACCTATTAATTAAGTTGTTATGATGATGATGATGATGATGATGATGATGAT 897
QY 168 -----PrometArgTyrPheAsnGlnGlyTyrProTyr-----AsnAsnLys 180
Db 898 TATCTTTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 957
QY 181 GluProLeuLeuProAsnGlnLysAsnHisAlaTyrThrGlyCysValAlaThrAlaAla 200
Db 958 -----ACCATTTGGAGACTTAAGCTTAAGAGCT----- 964
QY 201 GlnIleMetArgTyrHisSerThrProLeuGlnGlyGlyLysSerPheAspTyrHisAla 220
Db 985 -----TCATGTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1026

```

OY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetYrAspTrpIleAsnMet 240
 DB 1027 GSTAAC-----
 OY 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTrpAlaThrLeu 260
 DB 1033 ---GGTACATTAGATTAAATGGCGCAGTCAAAATCCCT---AATAAATGTGTCCACTA 1086
 OY 261 MetArgAspValSerAlaSerValSerMetSerPheTrpGluAsnGlySerGlyThrTrp 280
 DB 1087 -----CCTACGGCAGCTTCATTGCGGAGATAGGTCTTCCTCAATG--- 1125
 OY 281 SerValTrpValValGlyAlaLeuArgAsnAsnPheArgTrpLysArgSerLeuGlnLeu 300
 DB 1126 -----TTATTATTATGATGAAGATT----- 1146
 OY 301 HisValArgAlaLeuThrTrpSerGlnGluTrpHisAspMetIleArgGlyGluLeuAla 320
 DB 1147 -----AAGAAATGG-----TTATTAAATGGCGTGTACGT 1176
 OY 321 SerGlyArgProValTrpTrpAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCys 340
 DB 1177 GAAGGAATCCCTTATGCTGACAGTAGAAGACGATCAATTTACACGAAAAAGATTATTATT 1236
 OY 341 AspGlyTrpAlaSer---AspGlyThrPheHisPheAsnTrpGlyTrpGlyValSer 359
 DB 1237 CAAGGTATTCTTAATCAAGCATTTACAGCTAAATTT---TGGGAT-----ACT 1281
 OY 360 AsnGlyPheTrpLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGly 379
 DB 1282 AATGCTGAATTTACATTT-----AATATAGGAGTCAACCAATGCA 1323
 OY 380 IleGlyPheThrIleTrpGlnGluIle---IleThrGlyIleGluProAlaLysThrPro 398
 DB 1324 ACAGTGCACCAATCAAAAGTACATTAATAAAAGCTATTACAGCTGCAAGCATAGTG 1383
 OY 399 AlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLys---AspIleGluAla 416
 DB 1384 GGTCTTATGATTAATAGCCAACTCATGATGCTAGATGATAAATAGGCGATGAATCTCC 1443
 OY 417 GluTrpLysSerLeuSer---GlyLeuAsnValGlyTrpSerIleTrpAsn 432
 DB 1444 TCTTATTAAGGCTCAATTCATGTCGCGACATTTACATTCATGGAAGAAAGATTACTTT 1503
 OY 433 ThrGlyGluGlnSerAsnLeuAsnLeuGlyTrpArgLeuAsnLysAlaAspGlyGlu 452
 DB 1504 GCGCATCAAGGAACGAACTTTAAACATTTGAAATATATATATATCAAGTGCAGGTGCA 1563
 OY 453 Val-----IleGluValLysThrSerSerIleAsnIleSerTrpTrpGly 467
 DB 1564 TTGATTTTGAAGTAATTTTGTGTAAAGGCAATCAAAATATATATATCTTGCAAGGT 1623
 OY 468 TyrGly-----GluHisProGlySer 474
 DB 1624 GCAGCGCTTCTGTGAGAGAAAGTACTGTAATGCGAGGTGCATTAATCAAGAGCG 1683
 OY 475 PheSerLeuAlaPro----- 479
 DB 1684 GATCGCTTATCAAAATTTGGCTGGAGACCTTACTGTAATGCTAAAGGAAAAACTTA 1743
 OY 480 AsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTrpArgArgThrGlyThrGlu 1744
 DB 1744 GGAAGCTAGTGTGCTGAACGTTGGTTGTGTTA---GATCAACAAGCAGATGAA 1797
 OY 500 GlnTrpGluProValArgHisAlaGlnGlyTrpValAsn-----SerIleLys 516
 DB 1798 TCAGGCAAAAACAAAGCCCTTTAAAGAGTGTGCAATTTGATAGAGCTACCGTTCAA 1857
 OY 517 ValAsnThr-----ThrAspProAsnAsnValValVal-----Thr 528
 DB 1858 CTAATATAGCAGATCAAGTGTGATCTTAACAATATTATATTGCGCTTCTGCGTGGTGC 1917

OY 529 ValAspAsnAsnGlnGlyLysLeuSerIle-----ValProAsn-----Ser 542
 DB 1918 TTACATCTTAATGGGCAATTCATTAACCTTTGAAAGCATCAAAATACGAGTAAGCGCG 1977
 OY 543 PheValAlaAspLeuAsnSerTrpGlnHisSerThrIleThrVal-----Gln 558
 DB 1978 ATGATGTGCAAACCAACCGTCTCTCAACCGCAAAATATATAGCATTAAGCAGCAAGCAACT 2037
 OY 559 PheAsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGly 578
 DB 2038 ATTAATTCAGATTCACAAACACTT----- 2061
 OY 579 AlaThrAlaAspAspValIleSerLeuGlyTrpVal-----MetAlaGluVal 594
 DB 2062 ACTAATTAATAAAGATATTCATTTACGCGCTGTGTGTGTCAGCAAGATTAAGCTAAACAA 2121
 OY 595 ProGlyGly---SerSerAsnTrpProValValTrpSerLysAspValLeuThrLeuSer 613
 DB 2122 AATGCTCGTTAAATGATGATATATCAACAGTTPATGCAAGAAATCATTTGTTGCTTCT 2181
 OY 614 GluGlyAspTrpThrLeuTrpTrpArgPheSerIleAsnGlnLysAspGluTrpLys 633
 DB 2182 GGGGGG-----ACAAATTAACGCAATATACAG 2211
 OY 634 LysIleGlySerValSerVal-----LysThrProThrGluTrpThrHisPro 649
 DB 2212 CAATAATGTGTACGTTAGTTTATGATGTCGTGTCACCAAGCTCATGCTTACAAATCAT 2268
 OY 650 LeuPheGluValGlyHisAsnGlnThrSerThrTrpThrLeuAspMetAlaHisAsnArg 669
 DB 2269 -----TTAAGAAGCACTTCTCAACATGCA 2295
 OY 670 ValLeuPro-----AspPheThrLeuLysAsn 678
 DB 2296 GGTATCCCAACGCGCAATTTGTGGATCAGATTGATGATCAACCGCATTTAAAGCT 2355
 OY 679 LeuGlyLeuProPheAsnGlnGlyLeuValValPheArgGlnThrGlnSerSer 698
 DB 2356 GAAACCTTCAAATTAAGGCGGAGTGCAGGTGCTTCGCAATGTTCTTCAATTTGAG 2415
 OY 699 GlySerLeuTrpAlaAlaGlnGluThrValHisIleLysGlnGlyGlnThrPheValTrp 718
 DB 2416 GGAAT---TGG-----ACAGTCGCAATTAATCAATGCAATGCAATTT----- 2454
 OY 719 LysProValValGlyGlyProIleProAsp 728
 DB 2455 -----GGTGTGTGCCCAAT 2469
 RESULT 10
 US-10-160-758-4
 Sequence 4, Application US/10160758
 Publication No. US20030036076A1
 GENERAL INFORMATION:
 APPLICANT: EXELIXIS, INC.
 TITLE OF INVENTION: CADS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
 FILE REFERENCE: EX02-089C
 CURRENT APPLICATION NUMBER: US/10/160,758
 PRIOR FILING DATE: 2002-06-03
 PRIOR APPLICATION NUMBER: US 60/296,076
 PRIOR FILING DATE: 2001-06-05
 PRIOR APPLICATION NUMBER: US 60/328,605
 PRIOR FILING DATE: 2001-10-10
 PRIOR APPLICATION NUMBER: US 60/357,253
 NUMBER OF SEQ ID NOS: 16
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 4
 LENGTH: 10455
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-160-758-4

Alignment Scores:

Pred. No.: 0.00173 Length: 10455
 Score: 138.50 Matches: 199
 Percent Similarity: 33.90% Conservative: 127
 Best Local Similarity: 20.08% Mismatches: 336
 Query Match: 3.12% Indels: 32
 DB: 9 Gaps: 52

US-10-030-330-1 (1-843) x US-10-160-758-4 (1-10455)

QY 21 SerAlaProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnVal 40
 DB :|||:|||||
 DB 1420 AATGCCCGGAGTTCCACAGCTCCGAGTACAGCTGCGCATCATGAGTGCACAGCTC 1479
 QY 41 SerLeuArgMetGlyGln---ThrAlaValSerAspLysIleSerIleAspValTyr 59
 DB :|||:|||||
 DB 1480 GCGTTGGCCCTCCACTCTCTCCAGGTGGTGGACAG-----GlySer 75
 QY 60 ArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGlu-----GlySer 75
 DB :|||:|||||
 DB 1519 -----GATGAGACCCAGGCGCTGAACAGCATGTTGAGGTGATCTTGTTGGGGAGAC 1569
 QY 76 ProAlaTyrPheTyrValAlaAsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAla 95
 DB :|||:|||||
 DB 1570 AACTCCACACCATTCATCATCTCCCGACCTCGTCCAGGAGGAGCGGACATTCGTATT 1629
 QY 96 AspaParArgIleProThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAsp--- 114
 DB :|||:|||||
 DB 1630 CGGGGGCCCATCCCA-----CTGAGCTACGAGACCGGTGAGCCGCTTGTGATCTC 1683
 QY 115 -----SerMetProAspAsnLeu-----ArgMetIrpLeuGlnIleTyr 127
 DB :|||:|||||
 DB 1684 TTTCGCAATGAGAGTGGCTGACATGCGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 1743
 QY 128 AspGlnGluIleGlyLeuIleLeuSerGlyLysAlaGlnLeuAsnGlnIleLeuArg 147
 DB :|||:|||||
 DB 1744 AATGAAATGACACCGGCCCTCTCTGACACCGGACCATCTACACATCTGACGAG 1803
 QY 148 ThrGlnGluValProAlaGluValHisAlaLeuMet-----AspAsnGly----- 162
 DB :|||:|||||
 DB 1804 AACGTCACCGTGGGAGCTCTGTGCTGACAGTCTGCGACATGACATATGACAGCCACC 1863
 QY 163 -----HisPheAlaAsnAspProMetArgTyrPheAsnGlnIleTyrPro 176
 DB :|||:|||||
 DB 1864 TTTCGGAAGTCAGTACTTCTTCACTGATGACCGTGCAGAGGTTCGCTG----- 1914
 QY 177 TrpAsnAsnLysGluProLeuLeuProAsnGlnLysAlaTyrThrGlyCysValAla 196
 DB :|||:|||||
 DB 1915 -----GACAAAGAC-----ACGGGACTCATCATG 1938
 QY 197 ThrAlaAla-----AlaGlnIleMetArgTyrHisSerTyrProLeuGlnGly 212
 DB :|||:|||||
 DB 1939 CTGATTCGACAGCTGCTGACTGACGCTCATCCAGCCCTTCACCTGACATCTTCGCGG 1998
 QY 213 GluGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGlyThrPheGly 232
 DB :|||:|||||
 DB 1999 GACGGGGCGGGAGAGACACAGCGCGGTCCAGATCAAT----- 2040
 QY 223 GluMetTyrAspTyrIleAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSer 252
 DB :|||:|||||
 DB 2041 ---GTGTGGAT---GTCAAC-----GACAACTGGCCCACTTC 2073
 QY 253 GluValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMet----- 270
 DB :|||:|||||
 DB 2074 CAGAAAGATGCTTACGTGGGTCTCTGCGGAGAAAGAGACCTTCTGTCTCACAGCTGTG 2133
 QY 271 -----SerPheTyrGluAsnGlySerGlyThrTyrSerValTyr 283
 DB :|||:|||||
 DB 2134 CGGCTCGGGGCAACAGATGAACATCCCTCCCAACACCAATACATCAGCATTCG 2193
 QY 284 ValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeuHisValArg 303
 DB :|||:|||||
 DB 2194 AGTCATGTGCTTTCGAGAGCTTTCGACATC----- 2226

QY 304 AlaLeuTyrThrSerGlnGluTyrPheHisAspMetIleArgGlyGluLeuAlaSerGlyArg 323
 DB :|||:|||||
 DB 2227 AGCTGTACGAGGGCTAT-----GGATGTATCAGCGTCAGTCCG 2265
 QY 324 ProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyr 343
 DB :|||:|||||
 DB 2266 CCGCTCGATTATGAA----- 2280
 QY 344 AlaSerAspGlyThrPheHisPheAsnTyrPheGlyGlyValSerAsnGlyPheTyr 363
 DB :|||:|||||
 DB 2281 -----CAGATATCAATGCGGTGATT 2301
 QY 364 LysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGluGlyIleGlyPheThr 383
 DB :|||:|||||
 DB 2302 TATCTACCGGTATG-----GCCATGATGCTGGCAACCCCTCTCAACAGCACC 2352
 QY 384 IleTyrGlnGluIleIleThrGlyIleGluProAlaLysThrProAla----- 399
 DB :|||:|||||
 DB 2353 GTCCCTGTACACATCCAGGTGTTGATGAGATGACAAACCTCCACCTTCAGCAAGCCC 2412
 QY 400 -----GluAlaGlyThrAspAlaLeuProIleLeu 409
 DB :|||:|||||
 DB 2413 GCCTACTGCTCTCGGTGTGAGAACATCATGACAGGACCCAGGTGCTTCGTAAT 2472
 QY 410 AlaLeuLys---AspIleGluAlaGluTyrLysSerGlySerGlyLeuAsnValGlyTyr 428
 DB :|||:|||||
 DB 2473 GCCACAGACCTGACCGCTCCCGGAGTACGGCCAGAGATCC-----ATCATCTAC 2523
 QY 429 SerIleTyrAsnThrGlyGlnGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLys 448
 DB :|||:|||||
 DB 2524 TCCTTG-----GAAAGCTTCACCCAG-----TTTCGATCAATATGCC 2559
 QY 449 AlaAspGlyGluValIle-----GluValLysThrSerSerIle 461
 DB :|||:|||||
 DB 2560 CGCTCAGGGGAAATACACACACACAGCTCTGCTTACCAGGACCAACCACTGTAATCATC 2619
 QY 462 AsnIleSerTyrPyr-----GlyTyrGlyGluHisProGluSerPheSerLeuAla 478
 DB :|||:|||||
 DB 2620 CTTCATGCTTCGCGCATGAGCGGGGGTGGGCGACACCAAGAAACT----- 2667
 QY 479 ProAsnGlnLeuSerGlnGlyIleAsnThr-----IleThrLeuLeuTyrArgArgThr 496
 DB :|||:|||||
 DB 2668 -----GCGATCGCCACCGTAACATCACCTCTCGGACATCAATGAC 2709
 QY 497 GlyThrGlnGlnTyrProValArgHisAlaGlnGlyLysTyrValAsnSerIleLys 516
 DB :|||:|||||
 DB 2710 AACCAACCCACGAGGAGCAGCACCCTAC-----TACATCAACCTGTGTGAG 2757
 QY 517 ValAsnThrThrAspProAsn---AsnValValThrValAlaAspAsnAsnGlnGlyLys 535
 DB :|||:|||||
 DB 2758 ATGACCCCTCCAGACTCTGATGTGACACCGGTGGCTGTGACCAACCTGGGGAG 2817
 QY 536 -----LeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyr 550
 DB :|||:|||||
 DB 2818 AATGGACCTGTGTACAGCATCCAGCAACCAAGTTTACAGCCTCAACAGACACC 2877
 QY 551 -----GluHisSerThrIleThrValGlnPheAsnSerAspSerProAsp----- 565
 DB :|||:|||||
 DB 2878 ACGGGCAAGATCCGACACACCCAGCCATGCTGTGACCGGGAGAACCCGACCCCATGAG 2937
 QY 566 -----GluIleArgThrProValAlaPhe 573
 DB :|||:|||||
 DB 2938 GCGGACGTATGCGGCAAAATCGTCGTCTGTTACTGACTGTGGAGGCCCTCTGAAA 2997
 QY 574 AlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeu----- 587
 DB :|||:|||||
 DB 2998 GGCACACGACAGTCCACAGATGTTGTAACCTTGTGATCAATGACATGACACCCAC 3057
 QY 588 -----GlyTyrPheAlaGluValProGlyGlySerSerAsn 600
 DB :|||:|||||
 DB 3058 TTTCAGAACCTGCTTTTGTGGCGGAGGTGGAAGCATCCCGCGGGGGCTTCATC 3117
 QY 601 TyrProValValTyrSerLysAspValLeuThrLeuSerGluLysAspTyrThrLeuTyr 620

```

Db      3118 TACCAAGTGTG-----GCCATCGACCTCGATGAGGC----- 3150
OY      621 TTTATGPheserleAsnAsnGlnLysAspLutPrpLysLysleGlySerValSerVal 640
Db      3151 -----CTGAAC-----GCCGTGGTCTCTAC 3171
OY      641 LysThrProThrGluTyrThrHisProLeuPheGluValGlnHisAsnGlnThrSerThr 660
Db      3172 CGCATGCCGGGGGCGATGCCCCCGCATGAGCTCTCTCATCAACAGCAGCAGCGCGGTGGG 660
OY      661 TyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLys----- 677
Db      3232 CTCACCAACCAACCGAGCTGAGCCCGAGCGCGATCGCGAGTACCGCTCGCGGTGGG 696
OY      678 ---AsnLeuGlyLeuProPheAsnGlnGluLeuValValPheArgGlnThrGlnSer 696
Db      3292 AGTATCGACGACCGCC-----ACCAACCATGTCGTGGATGTAACAGCAG 3363
OY      697 SerSerGlySerLeuThrAlaAlaGlnGluThrValHisIle-----LysGln 712
Db      3319 TCCACGACGACGCTC-----ACCAACCATGTCGTGGATGTAACAGCAG 3363
OY      713 GlyGluThrPhe-----ValTyrLysProValValGluGlyProIlePro----- 727
Db      3364 AGCGCCACCTCTTCCCGCGCGCTGACATGTCTGTCTGCGAGAGCGTGGCAGCGGAG 3423
OY      728 -----AspGlySerTyrAlaThrLeuHis 736
Db      3424 TTCGGGTGTCTGTGCTGACAGCGACGACAGCGAGCGCGCCCATGCGAGCTCAGC 3483
OY      737 AlaPheValAsnGly-----GlnGlnGlnLeuTyrLeuGlyLysArgAsnTyrThr 754
Db      3484 TACTTCATCACAGGGGCAACGTGATGGAGAGTTCAGCTGGGTACCGCATGCCGTT 3543
OY      755 ValLysIleValAsnGly-----ThrAla-----ValGlu 764
Db      3544 GTGAGAACCGGTGGGGCTGTGACCGCGAGACCCAGCGCGCTCATCTCTGGAG 3603
OY      765 AlaIleGluSerSerGluGluIleArgValPheProAsnProAlaArgAspTyrVal--- 783
Db      3604 GCCATCGACAAAGCGCCCTGTAGGAGACGACAGCGGACAGCCAGCCGCTTCTCTACT 3663
OY      784 -----GluIleSerAla 787
Db      3664 GTCTGTGATGTGATACAAAGCGCCCATCTTCTGACAGAGCAGTATGAGCCAGCGCTC 3723
OY      788 Pro---CysIleProGlnGlnLutThrSerIleIleLeuPheAspLeuSer----- 802
Db      3724 CCGTGAAGACATCCCTGAAGGCGACAGCATCTTGCAGCTGAAAGCCAGCAGCATGAG 3783
OY      803 -----GlyLysIleValMetLysAsnSerLeuSerAlaGlnHisGly----- 816
Db      3784 GCGGAGTTTGGGCGT---GTGTGTACCGCATCTCTCATGTATGACATGGCAACACTTC 3840
OY      817 ArgMetAspValSer-----ArgLeuPro 824
Db      3841 CGGATCCATGTCAGCAATGGGCTCTGTATGCGAGGCGCGCGCCCTGTGACCGGAGCG 3900
OY      825 AsnGlyAlaTyrIleLeuLysValAspGlyTyr 835
Db      3901 AACTCATCCACGCTGCTGATAGTGAAGCGCTAC 3933

```

RESULT 11

US-10-030-624-5
Sequence 5, Application US/10090624
Patent No. US20020132335A1

GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOTO, Tomoko
APPLICANT: ASADA, Kiyoko
APPLICANT: KATO, Kunoshin

```

; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 4765
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; US-10-090-624-5

```

Alignment Scores:

```

Pred. No.: 0.000588 Length: 4765
Score: 138.00 Matches: 163
Percent Similarity: 31.13% Conservative: 87
Best Local Similarity: 20.30% Mismatches: 264
Query Match: 3.11% Indels: 290
Db: 12 Gaps: 42

```

US-10-030-330-1 (1-843) x US-10-090-624-5 (1-4765)

```

OY      185 ProAsnGlnValAsnHisAlaTyrThrGlnGlyCysValAlaIleThrAlaAlaGlnIleMetArg 204
Db      1174 CCTAACGAGAGATATGACGATATTGGG----- 1200
OY      205 TyrHisSerThrProLeuGlnGlyGluGlySerPheAspTyrHisAlaGlySerLeuVal 224
Db      1201 -----TGGATGTACAGCGTCAAGCAACT-----CAGGTAGCTGGAAGCTGTCT 1245
OY      225 Gly-----AsnTrp-----SerGly----- 229
Db      1246 GGTATGACAGCAACAAATGATGCTGGATGGCTGCATGATGTAAGTAAAGGAA 1305
OY      230 ThrPheGlyGluMetTyrAspThrPheIleAsnMetProGlnAsnProAspLeuAspAsnLeu 249
Db      1306 GTGTCTCAAGACTGATGTTGG-----GANTATACAGACGTT 1344
OY      250 ThrGln-----SerGlnValAspAlaTyrAlaThrLeu 260
Db      1345 ACCACAGACACCGCTGACGGGTGTGCTTCAGGTGCCAAATATGCAATAGAGTTCTT 1404
OY      261 MetArgAspValSerAlaSer----- 267
Db      1405 AGGAGTATGAGCGGGGTAGCATGTGGATATTATAGAGTATGACATACGACGCAAC 1464
OY      268 -----ValSerMetSerPhe-----TyrGlnAsnGlySer 277
Db      1465 CATGTGCAGACGTTAATACATGATGCTCGGTGAATATGCTTCATCTTAAATGATGTA 1524
OY      278 GlyThrTyrSerValTyrValValGlyAla-----LeuArgAsn 290
Db      1525 GATCCAGAAAGCGTGTGCTGTGATGA-CTTACCGAAAGTACGGGTGTGTTGTTGTAAT 1583
OY      291 AsnPheArgTyrLysArgSer----- 297
Db      1584 AGCTGCAGAAATGAAGGTCTCGCATTAACATCGTGGAGTCTCGTGTGTCACAA 1643
OY      298 -----LeuGlnLeuHisValArg-AlaLeuTyrThrSerGlnI 310
Db      1644 GGCATTAAGTGTGAGCTGTGCTGCACTGATTAACCTTGAGTATATGTTCCAA-- 1701
OY      310 utrpHisAspMetIleArgGlyGluLeuAlaSerGlyLysProValTyrTyr----- 327
Db      1702 -----GCACCTGGATATCCTGATTAAGTAAAGTATTA 1733
OY      328 -----AlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyrAl 344

```

```

Db      1734 TTACTCCCGCCCTACACAAAGCTTAGATAGCATTC---TTCTCAAGCAGAGGGCCGAG 1790
Qy      344 aserispglythrphnehispheasntpr-----glytrpglyglyvalseransgl 361
Db      1791 AATAGATGGTGAATAAAMACCAATGTAGTGGCTCCGATTAACGGA-----1836
Qy      361 yphetryllysleuthrleuSerProthrserleuuglyileglygluglyllegl 381
Db      1837 -----ATTACTCATCCCTGCCGATGTGGATTGGCGGA-----1869
Qy      381 ypherthriletyrgingluileierhrglyllegluProalalythProalaglual 401
Db      1870 -----GGGACTCATGTCTGGAACTTGATGGCT---ACCTCCATGTCCAG 1913
Qy      401 aglythrspalaleuProilleuualaleuylsaspilleglualeglutryllyser-- 420
Db      1914 CGGTGTCTGTGCATCCATAGCGGGGCAAGGCCGAGGAATTAATCTACATCCAGA 1973
Qy      421 -----gluserglyleuansvalglutryserlletyrAspthrhl 434
Db      1974 TTTAATTAAGAGGTTCTTGAGAGCGGTGCAACCTGGCTTGAGGGAGATCCATATACCTG 2033
Qy      434 yglugluglnSerAsnleuAspbleuglytyrarg-----445
Db      2034 GCAGAGTACTGACTGAGCTTGACCAAGGTCATGGTCTGTTAAGTTACCAAGTCTCGGA 2093
Qy      446 ----LeuanslysalaspglygluvalleluvallysthrSerSerleasnllese 464
Db      2094 AATCTTAAAGCTATTAACGCCACCACTCTCCCAATTTGATCACTGGGAGACAGCAACTG 2153
Qy      464 rtrptryglutrygluglnHisProgluserpheSerleualaproasnglnleusergl 484
Db      2154 CTACAGCCACTTGGCGGAGTAC-----TT 2177
Qy      484 ngllyleasnthrilethrleuLeutyrgargrghrglythrghlu-----glutry 501
Db      2178 GGGTGTGAGCATTATAGAGGTCTCTACCAAGAACTCTATACGTCGACATTTGTGAGTG 2237
Qy      501 pgluprovalarghisalaglnlyglutryvalasnserrileysvalasnthrthras 521
Db      2238 GCAC---ATAAGTACGTAGGGGACAGCGAGTACAGAACTTTGAACTATGCAACTGA 2294
Qy      521 pPro-----AsnansvalvalThrvAlaspsnAsnsglnugly 535
Db      2295 GCCATGAGATTAAAGCTTTTGTCAGTGGAGATGTAATTTAGAGAAACAATACGAGTTTGT 2354
Qy      2355 CCTTAGGGTG-----AAATATGATGTAGAGGGCTTGACCCAGGCTCTCA 2399
Db      555 ethrvalglnpheasnserserSerpaspglulearthyProvalAlaPheAlale 575
Qy      2400 TGTGGAAGGATATCATGTATGATGACCA-----ACACGCCAGTT-----2439
Db      575 userthrghlathrAlaspspvalilleserleuuglytrpvalmetAlagluvalpr 595
Qy      2440 -----ATTGAAGCAGAGATCTTG-----2457
Db      595 oglyglyserSeranslyrProvalAlatrPserlyasprvalleuthrleuserglugl 615
Qy      2458 -----AACACAAATGTATATCCGAGAA-----TTCACTCTGAGAA 2495
Db      615 yaspyrthrleu---trpyrargpheserilleasnsnglnlyaspsglutrpLysly 634
Qy      2496 CAATTACACCCCTCACCTGTAT-----2517
Db      634 slleglyserValserVallysthrProthrghlutytrhrhisProleuPhegluvalgl 654
Qy      2518 -----GATATTAAATGCTCCAGAAATGCGACTCACCTTCTTCACTGTGCC 2564
Qy      654 yHisasnglnthrSerthrtyrThrleuaspmetAlaHisasnargValleu-----Pr 672

```

```

Db      2565 TGAGAGGTGAGCGTCTCTACGCGATGACCATATCTAGGACTAGGCTGTACAGACC 2624
Qy      672 oAsp-----Pherthrleuylsasnleu-----gl 680
Db      2625 AGATGGAATGTTGTGTCCCATACAGCAATGATTAATCTCCGCTGACGTCAAAATCC 2684
Qy      680 yleuProPheasnglygluleuvalvalval-----PheargglnthrGlnserse 697
Db      2685 AATGCCGCAAACTGGAGCTAGTATGAGCTGATTAATCTTGACACCCCTCATATGACTC 2744
Qy      697 r-----SerglySerleuthrPalalaglngl 706
Db      2745 GGGCTTCCTGTAGAGATTACGAGATGAGATACCAAGGCTTGGTACATTAACAG 2804
Qy      706 uThr-----ValHisileysglnlygluthrPhevallytyrly 719
Db      2805 GACATACCTTGACACTAACACATGAAATTCATTAATTAATTAATTAATTAATTAATGC 2864
Qy      719 sProvalvalglugly-----ProileProaspGlySerTyrrArgAlathrleuHisal 737
Db      2865 CCCAATTAATGCACTCTAATCCCATTTGGCTTGACACTTACAAATCCGAGCTTGAAAG 2924
Qy      737 aphevalasnglyglnglnleuLeutyrglylysarargAsntrThrvalylsil 757
Db      2925 CGTT-----GGTATGAGAGAGTCTTCATTAAGGCG-----2955
Qy      757 eValasnglythrAlaValgluallegluserSerleuglnleuArgvalPheProas 777
Db      2956 -----ATTGAACTCTCTGAAAGCAGCCAGAGACTGAAAG-----2988
Qy      777 nProalArgAsptyrValglulleSerAlaProCysilePro---Glnluthrseril 796
Db      2989 -----ATTAGATAGGCAACCAACCAAGTTCCTCCAAATTCAGATCTAGACTT 3032
Qy      796 eileuPheaspLeuserelylysilevalmetlysasnserleuserAlaglyhisgl 816
Db      3033 GTACCTTTATGACAGTAAGGCAATTTAGTGGCTTAGAGTGAAGAACCAACAGAGAA 3092
Qy      816 yargmetaspvalaserArgleuProasnnglyAlaTyrlleuLysValaspglytyrth 836
Db      3093 AGAGGTGTAGTATGATGATATCCCTAAGCTGAGATTATTCATATGATGATGATGATCAG 3152
Qy      836 rThrlys 838
Db      3153 CCTCAGG 3159

RESULT 12
US-10-080-505-16
; Sequence 16, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RTT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; PRIORITY FILING DATE: 2002-02-22
; PRIORITY APPLICATION NUMBER: US 08/296,791
; PRIORITY FILING DATE: 1994-10-25
; PRIORITY APPLICATION NUMBER: US 09/839,996
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 4828
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS (4548)
; LOCATION: (313)..(4548)
; OTHER INFORMATION:
US-10-080-505-16
Alignment Scores:

```

Pred. No.: 0.0006
 Score: 138.00
 Percent Similarity: 31.708
 Best Local Similarity: 18.778
 Query Match: 3.118
 DB: 9
 Caps: 48

US-10-030-330-1 (1-843) x US-10-080-505-16 (1-4828)

1 MetLysLys-----SerPheLeuLeuAlaLeuValMetLeuPheGlyIle 15
 313 ATGAAAAAAGCTGATTTCTGCAATTTTAAACCGCTTCATTCATTA---GGATA 369
 16 AlaMetGln-----GlyHisSerAlaProValThrGlyGluArgAlaLeuSerLeu 32
 370 GATCCGACCGCTGGCGAGTCACTACT-----TATTTGGGATT 408
 33 AlaArgLeuAlaLeuArgGlnValSerLeuArgMetGlyGlnThrAlaValSerAspLys 52
 409 GACTACCAATATTTATGCTGATTTTCCGAGAAATGAAGGAAAGTTCCAGTGGGGCTAAA 468
 53 HisSerIleAspTyrValTyrArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGlu 72
 469 ---AAATTTGAGGTGTACAAATGAATGAATTTAGTTGCG---ACATCAATGACA 519
 73 GluGlySerProAlaTyrPheTyrValAlaAsnArgGlyAsnAsnGlnGlyTyrAlaLeu 92
 520 AAAGCCCAATGATTTGATTTTCCGCTGCTGCGGA-----AATGGGGTGGCGCA 570
 93 ValAlaAlaAspAspArgIleProThrIle-----LeuAlaTyrSerProIleGly 109
 571 TTGGTGGCGGATCAGTATATTTGATGATGCGACATATAGTACGCTTACCAATGTGGAT 630
 110 ArgPheAspMetLeuSerMetProAspAsnLeuArgMetIleProLeuGlnIle----- 126
 631 TTGTGCTGTAAGACAAATCCGATCAACATCGTTTACTTATTAATTTGGAACGG 690
 127 -----TyrAspGlnGluIleGlyLeuLeuLeuSer 136
 691 AATATATTATATAAAGCATCAACAGCATCTTATGAAAGACATC----- 735
 137 GlyLysAlaGlnLeuAsnGlnGluIleLeuArgThrGlyGlyValProAlaGluValHis 156
 736 CACAAACCCAGCTTACATTAATTTGTT-----ACGGAAGCCACCCCAATGATGACT 789
 157 AlaLeuMetAspAsnGlyHisPheAlaAsn-----AspProMetArgTyr 171
 790 TCTGATATGACGCAACAATAATACATAGACGAATAATCCCGAACCGGTGGTATTC 849
 172 AsnGlnGlyTyrPro---TyrAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAla 190
 850 GGCTCCGGGTGGCAGTTTGGCGAACAATCA-----AACCAAGCGCGAC----- 894
 191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTyrProLeu 210
 894 ----- 894
 211 GlnGlyGlnGlySerPheAspTyr-----HisAlaGlySerLeu 223
 895 CAAGTGGCGGCGCATATCATTTACCTACAGCAGCAGCAATACACAACAAGCGGAGCA 954
 224 ValGlyAsnTyrPheSerGlyThrPheGlyGluMetTyrAspTyrIleAsnMetProGlyAsn 243
 955 GGGGCGGCGTGCATACTGAGCGGCGAT-----GTGGCCCAAGCGGCAAT 1002
 244 ProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMetArgAsp 263
 1003 -----TACGGCCCATTT----- 1014
 264 ValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyrSerValTyr 283
 1015 -----CCTATTGCAAGCTCAACGCGGACAGCGGTGCGCTATGTT----- 1056

QY 284 ValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeuHisValArg 303
 Db 1056 ----- 1056
 QY 304 AlaLeuTyrThrSerGlnGluTyrPheHisAspMetIleArgGlyGluLeuAlaSerGlyArg 323
 Db 1057 ---ATTATATATGCGGCAAAAAAAGTGTGATTAAGCGGCTATTGAGCGGCAAC 1113
 QY 324 ProValTyrTyrAlaGlyAsnAsn-----GlnSerIleGlyHisAlaPheValCys 340
 Db 1114 CCT-----TGCGCGGAGACAGAGATACATTCACCTGTCGCAAGCTTTTTCAT 1167
 QY 341 AspGlyTyrAlaSerAspGlyThrPheHisPheAsnTyrPheGlyTyrPheGlyValSerAsn 360
 Db 1168 GAATGCTTGAAATAAGATTGGCTACATCGTT-----TAAAGCCATCGGCGAAT 1218
 QY 361 GlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGlyIle 380
 Db 1219 GGTGCATACACATTTACA-----GACAAAGCGGACGCGCAGC 1254
 QY 381 GlyPheThrIleTyrGln-----GluIleIleThrGlyIleGluProAla 395
 Db 1255 GGCATTTGCAACAACAAGACAGAGACCATCTGAGTCGACGCGCTTAAAGACGAC 1314
 QY 396 LysThrProAlaGluAlaGlyTyrAspAlaLeuProIleLeuAlaLeuLysAspIleGlu 415
 Db 1315 AAATTAACCTGCGGAAGCT-----AAAGACGATGTT 1344
 QY 416 AlaGluTyrLysSerGluSer-----GlyLeuAsnValGlyTyrSerIleTyr 431
 Db 1345 TACCAATTCACAAAGTCCAAATATATACCTGCTCGTGTGAATAGCGTGAACCTGTAT 1404
 QY 432 AsnThrGlyGlnGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGly 451
 Db 1405 TTCCGAGATCAAAAAAGCGCACCTTACTTTCACCAACATCAACACGCTGGCGGC 1464
 QY 452 -----GluValIleGluValLysThrSerSerIleAsnIleSerTyrGly 467
 Db 1465 GGTTCATTTTGAAGGTAACTTTACGATATCTTCAGAAATATAGCAACTGGCAAGT 1524
 QY 468 TyrGlyGluHisProGluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyLeuAsn 487
 Db 1525 GCTGAGTCGATGATGAGTGAACAGCT----- 1551
 QY 488 ThrIleThrLeuLeuTyrArgArgThrGlyThrGln-----Tyr 501
 Db 1552 ---ACTGTTACTTGGAAAGTAATATGTTGAAATGATCGCTTCTTAATAATCGGC 1605
 QY 502 GluProValArgHisAlaGlnGlyGlyTyrValAsnSerIleLysValAsnThrThrAsp 521
 Db 1606 AAAGCACATTTGCAGTTAAAGCCAAAGGGAATAAGGTTGATGACCGATAGCGGAT 1665
 QY 522 ProAsnAsnValValValThrValAspAsnAsnGlnGly-----Lys 535
 Db 1666 GGTAAAGTCATTTTGGACAGCAGCAGCATCAAGCAACAACAGCTTTAGTGAA 1725
 QY 536 LeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIle 555
 Db 1726 ATTGGCTGTGTAGTGAGAGAGTACGTTCACTTAACGATGACAG----- 1773
 QY 556 ThrValGlnPheAsnSerAsp----- 562
 Db 1774 ---CAATTTATCTGATAAATTTATTTGCGCTCCGTGCTGCTTACATCTT 1827
 QY 563 -----SerProAspGluIleArgThrProVal 571
 Db 1828 AATGGCATTCATTAACCTTTAAAGTATTCACAAATACGATGAGAGACAGATGTT 1887
 QY 572 AlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTyrPheMet 591
 Db 1888 AATCACAATGCCCAACAGAAATCTACAGTG----- 1917
 QY 592 AlaGluValProGlyLysSerSerAsnTyrProValValTyrSerLysAspValLeuThr 611


```

Db      1918 -----ACATTACGTGCAGCGATTACATTAAAT 1944
Oy      612 LeuSerGluGlyAspTyrThrLeuTyrPheSerIleAsn-----AsnGln 628
Db      1945 GACAAACACTGGCGATTAAACGATATGATGTTTAAATGTTGTTGGTGCAT 2004
Oy      629 LysAspGluTyrPheLysIleGly-----SerValSerValLysThrProThrGluTyr 646
Db      2005 AAAGATGATCTAAATAATCTGAGCGTTGAAATGTTACTTACATCCGCTTAACAAGAT 2064
Oy      647 ThrHisProLeuPheGluValGlnHisAsnGlnThrSerThrTyrThrLeuAsp----- 664
Db      2065 AATCACTTCCTCTCATCAGGTGGAACAATTTAAAGCAATTTTACTCAAGCGGTGGC 2124
Oy      665 -----MetAlaHisAsnArgValLeuProAspPhe 674
Db      2125 ACTTAACTGTTTAAAGTGTGCGCCACACACGCAATTCATCATTTA-----AAT 2175
Oy      675 ThrLeuLysAsnLeuGlyLeuProPheAsnGlyGluLeuValVal----- 689
Db      2176 CCGCTTAACGACTTGGCGGACCT---AAGGCGAAGGTGTTATTCATGACGATTGGATC 2232
Oy      690 -----ValPheArgGlnThrGlnSerSerGlySerLeuTyrPheAlaGlnGluTyr 707
Db      2233 AACCGTACATTTAAA-----GCTGAAGAAC 2256
Oy      708 ValHisIleLysGlnGlyGluThrPheValTyrLysProVal-----ValGluGlyPro 725
Db      2257 TTCCAAATTTAAAGGCGAGTGGCGGTTCCTCCAAATGTTCTTCAATTGACGAAT 2316
Oy      726 ---IleProAspGlySerTyrArgAlaThrLeuHisAlaPheValAsnGlyGlnGln 744
Db      2317 TGGACAATCAGCAATTAACCCACACGACATTTGGTGTGCGCAATCAACAAATACC 2376
Oy      745 LeuTyrLeuLysGlyLysArgAsn-----TyrThrValLysIleValAsnGlyTyrAla 762
Db      2377 ATTCCAGCGCTTCAGATTTGACAGAGATTACGACTTTGAAACATTTAATTAAACGAT 2436
Oy      763 ValGluAlaIleGlnSerSerGluGluIleArgValPheProAsnProAlaArgAspTyr 782
Db      2437 AAAAAAGATTATGATTC----- 2454
Oy      783 ValGluIleSerAlaProCysIleProGlnGluThrSerIleIleLeuPheAspLeuSer 802
Db      2455 -----ATACCGACAACA-----CAATTAAT 2475
Oy      803 GlyLysIleValMetLysAsnSerLeuSerAlaGly---HisGlyArgMetAspValSer 821
Db      2476 GCGCTAATTAATTAATACTAATATGCAACAGTAATATTCATGCT----- 2520
Oy      822 ArgLeuProAsnGlyAlaTyrIleLeuLysValAspGlyTyrThrLysIleAsn 840
Db      2521 -----TTAGCAAACTTAAGTATGCTCACTTAATTAAT 2556

```

RESULT 13

US-09-848-616-169

Publication 169, Application US/09848616

Publication No. US20030054010A1

GENERAL INFORMATION:

APPLICANT: Sebbel, Peter

APPLICANT: Dunant, Nicolas

APPLICANT: Bachmann, Martin

APPLICANT: Tisot, Alain

APPLICANT: Lechner, Franziska

TITLE OF INVENTION: Molecular Antigen Array

FILE REFERENCE: 1700.0180002

CURRENT APPLICATION NUMBER: US/09/848,616

CURRENT FILING DATE: 2001-05-05

NUMBER OF SEQ ID NOS: 186

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 169

LENGTH: 8464

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PFMAICDFG
US-09-848-616-169

Alignment Scores:
Pred. No.: 0.00141
Score: 138.00
Percent Similarity: 31.36%
Best Local Similarity: 17.72%
Query Match: 3.11%
DB: 9
Matches: 8464
Conservative: 141
Mismatches: 307
Indels: 403
Gaps: 53

US-10-030-330-1 (1-843) x US-09-848-616-169 (1-8464)
Oy      56 AspTyrValTyrArgGlnGlyAspAlaGluArgGlyLeuThrSerGlnGluGlySer 75
Db      4535 GATGTCACATTTAATACGGCGCAGTGAACAGGAGTTCCTGCTGCACACGCGCG 4594
Oy      76 ProAlaTyrPheTyrValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuValAlaAla 95
Db      4595 CAACCTGCCAGTATGGGCGGTGAATACGCGCTTGTCCGCGTATGAATCTGCGCGGAT 4654
Oy      96 AspAspArgIleProThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSer 115
Db      4655 GATGCTGTGTGCA-----TTAACCACA 4678
Oy      116 MetProAspAsnLeuArgMetTyrPheGlnIleTyrAspGlnGluIleGlyLeuIleu 135
Db      4679 ATGTCACAGACCTACTGCGCATCTGCATCTGTGTCAGCAGGACTGAACCTGACGATC 4738
Oy      136 SerGlyLysAlaGlnLeuAsnGluGluIleLeuArgThrGluGly---ValProAlaGlu 154
Db      4739 CCTCAGCAGATTTATGAGTAT-----CGCCGCGGTGTTATATTCCTCCGAC 4786
Oy      155 -----ValHisAlaLeuMetAspAsnGlyHisPheAlaAsnAspPromet 169
Db      4787 TTATGGGATCCCGTATTAATGCCGATTCCTCAATTATTAATTCACGGAAATAGTGA 4846
Oy      170 ArgTyrAsnGlnGlyTyrProTyrAsnAsnLysGluProLeuLeuProAsnGlyAsnHis 189
Db      4847 CACAAACGCAATGCGG-----GTAACACGCAATAT 4876
Oy      190 AlaTyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTyrPro 209
Db      4877 GCATATTTAACCTACAGACTGG-----TTAATATTTGGTGGCGT 4921
Oy      210 LeuGlnGlyGluLysSerPheAspTyrHis-----AlaGlySerLeu--- 223
Db      4922 TTACCGGACAAATACCACTGGAGTTATACAGTAGCAGCAATCATCAGTAGCAAAAT 4981
Oy      224 -----ValGlyAsnTyr----- 227
Db      4982 AAATGCGACAGATATACCTGCTTGACGAGACATAATACCGTTACCTCCGCGCTG 5041
Oy      228 -----SerGlyThrPheGlyGluMetTyrAspTyrIleAsnMetProGlyAsn 243
Db      5042 ACCTGGGTGATGATTATACAGGCGGATATTTTGCATGTTTAACTTTCGCGCGCA 5101
Oy      244 -----ProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMet 261
Db      5102 CAATTGGCTCAGATGACAAATATGTTACCGGATAGTCAAAGAGAGATTGCGCGGTATC 5161
Oy      262 ArgAspValSer---AlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
Db      5162 CACGGATATGCTGCTGATCTGCACAGGTACATTAACAAATAGGATGACATTTAT 5221
Oy      281 Ser-----ValTyrValValGlyAlaLeu 288
Db      5222 AATAGTACGATGCGACCGGCGCTTTTACCATCAACGATATATTCGCGAGGT----- 5275
Oy      289 ArgAsnAsnPheArgTyrLysArgSerLeuGlnLeuHisValArgAlaLeu----- 305

```

Db 5276 -----ATATGCTGACTTGCAGGTACGATCAAGAGGCTGACGACG 5320
OY 306 -----TyrThSer-----GlnGluTrpHisasp 313
Db 5321 ACGGAGATTTTACCGTACCTATTCCTCAGTCCCGCTTTCAGACGTGAAGGCGAT--- 5377
OY 314 MetLeaArgGluLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSer 333
Db 5378 ---ACTGTTATTCATTCAGTCAGGAGAA---TACCGTAGTGAATGCGCAGAG 5428
OY 333 -----
Db 5429 GAAAAAACCCGCTTTCAGAGTACATTACTCAGCGGCTTCGCGTCCGTCGACAAATA 5488
OY 334 TleGlyHisAlaPheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrp 353
Db 5489 TATGTGGAACGCACTGGCGGATGCTTTCGTGCT---TTTAATTTC 5533
OY 354 GlyTrpGlyValSerAsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeu 373
Db 5534 GGTATCGG-----AAAACATG 5551
OY 374 Gly---TleGlyGlyGluGlyTleGlyPheThrTleTyrGlnGluTleThrGlyTle 392
Db 5552 GGGGCACTGGCGCTCTGCTGTGATGACGCGCTAATTCACACTT--- 5602
OY 393 GluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuProAlaLeuAlaLys 412
Db 5603 -----CCCGATGACAGTCAGCATGACGACAGACATCGGCTTTCCTCTAT 5647
OY 413 AspIleGluAlaGluTyrLysSerGlySerGlyLeuAsn-----ValGlyTyr--- 428
Db 5648 AAC-----AAATGCGTCAATGATCAGCAGCAATTTTCAGTACGTAGCGTTACCGT 5698
OY 429 -----SerIleTyrAsnThrGlyGluGlnLysSerAsnLeuAspLeuGlyTyr 444
Db 5699 TATTCAGACAGCGGATTTTAAATTCGCTGATACAAACATACAGTCGATGAATGAGCTAC 5758
OY 445 ArgLeuAsnLysAlaAspGlyGluValIleGluValLys--- 457
Db 5759 AACATGGAAACACAGACGGA---GTTATTCAGGTTAAAGCGAAATTCACGCACTATAC 5815
OY 457 -----
Db 5816 AACCTGGCTTATTAACACGCGGAATTAACACTCACCCTTACTCAGCACTGGCGGC 5875
OY 458 ThrSerSerIleAsnIleSer-----TrpTyrGlyTyrGlyGlnHisPro 472
Db 5876 ACATCAACACTGATTTGAGTGTAGCATCAAACTTATTTGGGAACAGATATGTCGAT 5935
OY 473 GluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeu 492
Db 5936 GAGCAATTCAGAGCTGATTAATAATGCTTCGAGATATCAACTG---ACGCTCAGC 5992
OY 493 Tyr-----
Db 5993 TATACCTGACGAAAAACGCGCTGCAAAAAGAGCGGATGATGTTAGCGCTTAACGTC 6052
OY 494 -----ArgArgThrGlyThrGlnGluTrpGluProValArg 505
Db 6053 AATATTCCTTTCAAGCCACTGCGCTGCTGACAGTAATAATCTCAGTG---GCA 6103
OY 506 HisAlaGlnGlyTyrValAsnSerIleLysValAsn----- 518
Db 6104 CATGCCAGTCCAGCTACAGCATGTCACAGATCTCAAGGTCGATGACCAATCTGCT 6163
OY 519 -----ThrThrAspProAsnAsnValIleValIleThrVal--- 529
Db 6164 GGTGTATACGCTACGTTGCTGGAAGACAACTCAGCTATAGCGTCAACCGCGCTAT 6223
OY 530 -----AspAsnAsnGlnGlyLysLeuSerIleValProAsnSerPheValAla 545
Db 545 -----
Db 6224 GCCGGGAGGCGATGGAATACGGA-----AGTACAGCTACGCC 6265
OY 546 AspleuAsn-----SerTyrGlnHisSerThrIle 555
Db 6266 ACGGTAAATTAATGCGGTGTTACGGCAATGCCAATATGCGTTACAGCATAC--- 6319
OY 556 ThrValGlnPheAsnSerAspSerProAspGlnIleArgThrProValAlaPheAlaLeu 575
Db 6320 -----GATGATATTAG---CACTCTATTACGAGATC 6349
OY 576 SerThrGlyAlaLeuThrAlaAsp---AspValIleSerLeuGlyTyr----- 589
Db 6350 ACGGCGGTGCTACGTGCTATCCCAATGCGTAACGTCGGGCAAGCCGTTAAACATACG 6409
OY 590 ---ValMetAlaGluValProGlyLysSerSerAsnTyrProValIleTrpSerLysAsp 608
Db 6410 GTGTCGCTTGTAAAGCGCTGCG---GCAAAAGAT 6442
OY 609 ValLeuThrLeuSerGluGlyAspTyrThrLeuTrpTyrArgPheSerIleAsnAsnGln 628
Db 6443 GCA-----AAAGTCGAAACCAAG 6460
OY 629 -----LysAspGluTrpLysLysIleGlySerValSerValLysThrProThrGlu 645
Db 6461 ACGGGGTGCGTACCGACACTGCGT---GCTATGCGGTGCTGCTTATGCCACTGAA 6514
OY 646 TyrThrHisProLeuPheGluValGlnHisAsnGlnThrSerThrTyrThrLeuAspMet 665
Db 6515 TATCGG-----GAAATATAGATGGCGGTGATACCAATACCTG 6553
OY 666 AlaHisAsn----- 668
Db 6554 GCTGATTAACGTCATTTAGATACGGGCTTGCATACGTTGCCACTGTCGGCGATC 6613
OY 669 -----ArgValLeuProAspPheThrLeuLys 677
Db 6614 GTGCGACAGATTTAAACGCGCGCTTGGATTAACATGCTCATACGCTACCCACACAT 6673
OY 678 AspLeuGlyLeuProPheAsnGlnGlyLeuValValValPheValGlnThrGlnSerSer 697
Db 6674 AATTAACCCGCTGCGCTTGG---GGAATGTCACATGAGAGATGACAGAT 6724
OY 698 SerGlySerLeu----- 701
Db 6725 ACGGCGATTTGCGGATTAATGTCAGCTTAACTCAGACGGAATGCTTAAAGCGGAAA 6784
OY 702 -----TrpAlaAlaGlnGluThrValHisIle----- 710
Db 6785 GTTCAGGTGAATGGGAGAGAGAAATGCTAGTGTGCCAATTAATCACTGCCA 6844
OY 711 LysGlnGlyGluThrPheValTyrLysProValAla-----GlnGlyPro 725
Db 6845 CCAAGAGCTCAGCAGCA---GTTATTAACCCAGCATCAGTGAATGTCGTAAGGGGCGT 6903
OY 726 IleProAspGlySerTyrArgAlaThrPheHisAlaPheValAsnGlyGlnGlnLeu 745
Db 6904 GATGAGAAACAACCTTTTATCTGTGTCGCTTGTGGCTGCGCGAGTCAAGC 6963
OY 746 TyrLeuLysGlyLysArgAsnTyrThrValLysIleVal-----AsnGlyThrAlaVal 763
Db 6964 TTGGCTGCGGATGACAGCATTAATCCGCGGTATGTCAGGATTAACGGCTGTAAGT 7023
OY 763 LglValIleGluSerSerGluGluIleArgValPheProAsnProAlaIleAspTyrVal 783
Db 7024 GGCGCGTGAATCAACCAATTTACGTGATGTCATGGAACCGCGGAGCAATTTAA 7083
OY 783 LglIleSerAla-----ProCysIlePro-----Gly 792
Db 7084 CAACATTCGCGGAGCACTCTGTTTCATTTGCTATTTTGTGTCACCCGTGGTAA 7143
OY 792 nGluThrSerIleIleLeuPheAspLeuSerGly-----Lys 804
Db 7144 TCCGCTTCTCGCTAAAGTGTGGTTACTGCGCTTGCAGATAGCCCAATGCCAATCT 7203

```
QY 804 silevalmetlysasnserleuseralaglyhisgly 816
DB 7204 GCTTCACCTTGAATAATACGGTGTGACGGCTTCGGGA 7240

RESULT 14
US-09-848-616-168
; Sequence 168, Application US/09848616
; Publication No. US20030054010A1
; GENERAL INFORMATION:
; APPLICANT: Sebbel, Peter
; APPLICANT: Dumant, Nicolas
; APPLICANT: Bachmann, Martin
; APPLICANT: Tisot, Alain
; APPLICANT: Lechner, Franziska
; TITLE OF INVENTION: Molecular Antigen Array
; FILE REFERENCE: 1700.0180002
; CURRENT APPLICATION NUMBER: US/09/848,616
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: PatentIn Ver. 2.1
; EQ ID NO 168
; LENGTH: 9299
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PRIMAIDCFGH
US-09-848-616-168

Alignment Scores:
Score: 0.00163 Length: 9299
Percent Similarity: 138.00 Matches: 183
Best Local Similarity: 31.358 Conservative: 141
Query Match: 17.728 Mismatches: 307
DB: 9 Indels: 403 Gaps: 53

US-10-030-330-1 (1-843) x US-09-848-616-168 (1-9299)

QY 56 AspyrValTYrArgInGlyAspAlaGluArgGlyIleThrSerGlnGluGlySer 75
DB 4535 GATGCACATTTAAATACGGCGCACAGTGAACAAGGATTTGTCCTGCCTGACACCGCGG 4594
QY 76 ProAlaTyRheTyValAlaAsnArgGlyAsnAsnGlnGlyTyRAlaLeuValAla 95
DB 4555 CAACGCCGACATATGGGCGGTGATACGGCTTGTGCGCGGTATGATGATGCGCGGAT 4654
QY 96 AspaParGileProThrIleLeuAlaTySerProIleGlyArgPheAspMetAspSer 115
DB 4655 GATCGCTGTGTCGA-----TTAACCCACA 4678
QY 116 MetProAspAsnLeuAlaArgMetIlePleuGlnIleTyRAspGlnGluIleGlyLeuIleu 135
DB 4679 ATGGTCCAGACGACATCGCATCTGATGTGTGTGACAGCGACGACTGAACCTGACGATC 4738
QY 136 SerGlyLysAlaGlnLeuAsnGlnGluIleLeuArgThrGlnGly---ValProAlaGlu 154
DB 4739 CCTCAGGCAATTTATAGTAAT-----CGGGCGCGGTGATTATTCCTCTCTGAG 4786
QY 155 -----ValHisAlaLeuMetAspAsnGlyHisPheAlaAsnAspPromet 169
DB 4787 TTATGGATCCCGGATTAATATCGCGATTCGCAATTAATTTTCAGCGGAATATGTGA 4846
QY 170 ArgTrpAsnGlnGlyTyRProThrAsnAsnLysGluProLeuLeuProAsnGlnLysHis 189
DB 4847 CAGAAATCGAATGGG-----GCTAACACCATATAT 4876
QY 190 AlaTyRThrGlnGlyCysValAlaThrAlaAlaGlnIleMetArgTyRHisSerTrpPro 209
DB 4877 GCATATTTAAACCTACAGAGTGG-----TTAAATATGTCGTCGCGCT 4921
QY 210 LeuGlnGlyGluGlySerPheAspTyRHis-----AlaGlySerLeu--- 223
DB 4922 TTACGGCAATATCCACCTGGAGTTATTAACATGACGACATCATCATGACCAAAAT 4981
QY 224 -----ValGlyAsnTrp----- 227
DB 4982 AAATGACAGATATATACCTGCTTGACGACATTAATATACCGTTACGTCGCCGCTG 5041
QY 228 -----SerIleThrPheGlyGlnMetTyRAspTrpIleAsnMetProGlyAsn 243
DB 5042 ACAGTGGGATGATGATATACCTGAGGATATTTGATGATGATTTACCTTCCGCGCGCA 5101
QY 244 -----ProAspLeuAsnLeuThrGlnSerGlnValAlaPheAlaThrLeuMet 261
DB 5102 CAATTGGCTCAGATATACATATATGTCACCGATACGATACGAAGAGATTTGCCCGGATC 5161
QY 262 ArgAspValSer---AlaSerValSerMetSerPheTyRGlulAsnGlySerGlyThrTyR 280
DB 5162 CAGCGATATGCTGCTGCTACTGCACAGCATCATTTAAACAAATGAGTATGACATTTAT 5221
QY 281 Ser-----ValTyRValAlaGlyAlaLeu 288
DB 5222 AATAGTACGGTCCACCGGGCTTTTACCATGACAGATATCTATGCGCGAGGT----- 5275
QY 289 ArgAsnAsnPheArgTyRlysArgSerIleuGlnLeuHisValAlaArgAlaLeu----- 305
DB 5276 -----AATAGTGTGACTGTACAGTAAACATCAAGAGCGCTGACGCGCAGC 5320
QY 306 -----TyrThrSer-----GlnIleTrpHisAsp 313
DB 5321 ACAGATATTTTACCTACCCCTATGCTGCTAGTCCGCTTTTGCACAGTGAAGGCGAT--- 5377
QY 314 MetIleArgGlyLysLeuAlaSerGlyArgProValTyRTrpAlaGlyAsnAsnGlnSer 333
DB 5378 ---ACTGTTATTCATTTACGGCAGAGAA-----TACCGTAGTGAAATGCGCAGCAG 5428
QY 333 ----- 333
DB 5429 GAAAAAACCCGCTTTTCCAGATACATATCTCCACGCGCTTCGCGCTGCGACATA 5488
QY 334 IleGlyHisAlaPheValLysAspGlyTyRAlaSerAspGlyThrPheHisPheAsnTrp 353
DB 5489 TATGGTGAACGACGACGCTGCGATGCTATCGTCT-----TTTAATTTTC 5533
QY 354 GlyTrpGlyGlyValSerAsnGlnPheTyRlysLeuThrLeuLeuSerProThrSerLeu 373
DB 5534 GGTATCGG----- 5551
QY 374 Gly---IleGlyGlyGlnGlyIleGlyPheThrIleTyRGlulIleIleThrGlyIle 392
DB 5552 GGGCGACGCGCGCTGCTGCTGATATGACGACGAGCTAATTCACACTT----- 5602
QY 393 GluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLys 412
DB 5603 -----CCGATGACAGTCAAGATGACGAGACATCGGCTTTTCTCTAT 5647
QY 413 AspIleGluAlaGluTyRlysSerGlyLeuAsn-----ValGlyTyR--- 428
DB 5648 AAC-----AAATCGCTCAATGATATGACGACGATATTTAGTTAGGTGATTCAGT 5698
QY 429 -----SerIleTyRAsnThrGlyGlnGluGlnIleSerAsnLeuAspLeuGlyTyR 444
DB 5699 TATTCGACACGCGATATTTATTTTCGCGATACAAACATACAGTGAATGATGATGCTAC 5758
QY 445 ArgLeuAsnLysAlaAspGlyGluValIleGluValLys----- 457
DB 5759 AACATTGAACACAGACGCGA---GTTATTCAGAGTTAAGCGGAATTCACCGCATATTAC 5815
QY 457 ----- 457
DB 5816 AACCTCGCTTAATACAAACGCGGAAATTAACAATCAACGCTTACTACGACAACTCGGCGC 5875
QY 458 ThrSerSerIleAsnIleSer-----TrpTyRGlulTyRGlulHisPro 472
DB 5876 ACATCAACACTGATTTTGAAGTGTACCATCAAACTTATTTGGGAACGACATATGTCGAT 5935
```


OY 626 AaAaNgInLyAspGluTIPlyslIleGlySerValSerValIlyThrProthGlu 645
Db 277 -----CAGATGAAACAAACAATCCGCGGCTGGATATGTGACGATTAAGTCAAG 227
OY 646 TTTTTHrHSPProLeuPheGluValGlyHisAsnGlnThrSerThrThrLeuAspMet 665
Db 226 TTC-----AACACAGCGGCAACTCATCTTGTATTAAC 194
OY 666 AAlHAsnArGValLeuProAspPheThrLeuLysAsnLeuGlyLeuProPheAsnGly 685
Db 193 AGCAAAAGCCGCTTCCTTCACAGGCGCAATGCTAACGCCAGTTATATAACACCGGT 134
OY 686 GluLeuValValPheArgGlnThrGlnSerSerGlySerLeuThrAlaAlaGln 705
Db 133 GAAATGACGCTG-----GATCGCGCATTTAGGTGCGGC--- 101
OY 706 GluThrValHISleIleGlyGlnGlyLeuThr-----PheValTyrLysProValValGlyGly 724
Db 100 ---GCTATTCACTAACGACGATGCGCGCAGTTTATTAAAGGCGGTGGATGCG 44
OY 725 ProIle 726
Db 43 AAGATC 38

RESULT 16

US-10-160-758-1
Sequence 1, Application US/10160758
Publication No. US20030036076A1
GENERAL INFORMATION:
APPLICANT: EXELIXIS, INC.
TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
FILE REFERENCE: EX02-089C
CURRENT APPLICATION NUMBER: US/10/160,758
PRIOR FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US 60/296,076
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/328,605
PRIOR FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: US 60/357,253
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 10065
TYPE: DNA
ORGANISM: Homo sapiens
US-10-160-758-1

Alignment Scores:

Seq. No.: 0.00265 Length: 10065
Percent Similarity: 136.50 Matches: 199
Best Local Similarity: 32.80% Conservative: 126
Query Match: 20.08% Mismatches: 337
DB: 9 Gaps: 52
Indels: 329

US-10-030-330-1 (1-843) x US-10-160-758-1 (1-10065)

OY 21 SerIaSProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnVal 40
Db 1030 AATCCCGGAGAGTTCAACAGCTCCGAGTACAGCTGCGCATCTGAGCTGCACAGGTC 1089
OY 41 SerLeuArgMetGlyGln---ThrAlaValSerAspLysIleSerIleAspTyrValLys 59
Db 1090 GGGTTTGCCCTTCCACTCTTCATCCAGGTGGTGACAG----- 1128
OY 60 ArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGluGlu-----GlySer 75
Db 1129 -----GATGAGAATTGGCGCTGAACAGCATGTTTGAAGTACTTGGTGGGAC 1179
OY 76 ProAlaTyrPheTyrValAlaAsnArgGlyLysAsnGlnGlyTyrTrAlaLeuValAlaAla 95

Db 1180 AACTCCCAACACATTCATCATCTCCCGACCTCCGTCAGGGGAAGCGGACATTCGTAT 1239
OY 96 AspAspArgTyrIleProThrIleLeuAlaTyrSerProIleGlyLysPheAspMetAsp--- 114
Db 1240 CGGGTGGCCATCCCA-----CTGACCTAGAGACCGGTGACCGTACGATTTGTCTC 1293
OY 115 -----SerMetProAspAsnLeu-----ArgMetTyrLeuGlnIleTyr 127
Db 1294 TTGGCCAAATGAGAGTGTGCTGACACATGTGGCTATGCCAAGGTGAGATCATCTCATC 1353
OY 128 AspGlnGlnIleGlyLeuIleLeuSerGlyLysAlaGlnLeuAsnGlnGlnIleLeuArg 147
Db 1354 AATGAAATAGAACACCGGCCATCTTCAGCCGACAGCTGCAACATCATGAGCTGTACAG 1413
OY 148 ThrGlnGlyValProAlaGluValAlaHisAlaLeuMet-----AspAsnGly----- 162
Db 1414 AACGTACACCGTGGGACCTGTGCTGACAGCTCTGGCAACTGCACATGATGACGAGCACC 1473
OY 163 -----HisPheAlaAsnAspProMetArgTyrAsnGlnGlyTyrPro 176
Db 1474 TTGGGGAAGTCAGCTACTTCTTCATGATGACCTGACAGGTTCCTGCTG----- 1524
OY 177 TrpAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAla 196
Db 1525 -----GACAAGAC-----ACGGGACTCATATG 1548
OY 197 ThrAlaAla-----AlaGlnIleMetArgTyrHisSerTyrProLeuGlnGly 212
Db 1549 CTGATTCGACAGCGCGGACGTATGAGCTATCCACCGCTTCACCTGCATCATTCGCCGG 1608
OY 213 GlnGlySerPheAspTyrHisAlaGlySerLeuValGlyAspTyrSerGlyThrPheGly 232
Db 1609 GACGGGGCGCGGACGAGACACACAGCGCGGTACAGATCAT----- 1650
OY 233 GluMetTyrAspTyrPheAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSer 252
Db 1651 ---GTGTGGAT---GTCAAC-----GACAAGTCCCACTTC 1683
OY 253 GlnValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMet----- 270
Db 1684 CAAAGAGATGCTTACGTGGTGTCTGCGGAGAACAGCTTGTGTCACACAGCTGTG 1743
OY 271 -----SerPheTyrLysAsnGlySerGlyThrTyrSerValTyr 283
Db 1744 CGGTCGCGGCAACGATGAGATCCCTCCCAACACAGATACAGTACAGTATTC 1803
OY 284 ValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnHisValArg 303
Db 1804 AGTCATCTGCTCTTGGACAGCTACTTCGACATC----- 1836
OY 304 AlaLeuTyrThrSerGlnGlnGluTyrHisAspMetIleArgGlyLeuAlaSerGlyArg 323
Db 1837 AGCCTGACGAGGCTAT-----GGAGTATCAGCTCAGTGC 1875
OY 324 ProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspLysTyr 343
Db 1876 CCCCTGGATTATGAA----- 1890
OY 344 AlaSerAspLysThrPheHisPheAsnTyrGlyTyrGlyValSerAsnGlyPheTyr 363
Db 1891 -----CAGAAATCCAAATGGGCTGAT 1911
OY 364 LysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGlnGlyIleGlyPheThr 383
Db 1912 TATCTGACGGTATG-----GCCATGATGCTGCAACCCCTCTCAACAGCACC 1962
OY 384 IleTyrGlnGluIleIleThrGlyIleGluProAlaLysThrProAla----- 399
Db 1963 GTCCCTGCACACATGAGAGTGTGATGAGAAAGACAACCTCCACCTTCACAGAGCC 2022
OY 400 -----GluAlaGlyThrAspAlaLeuProIleLeu 409
Db 2023 GCTACTTCTGCTCTGCTGAGAAACATGATGCAAGAGCCACGGTGTCTTCTGAT 2082


```

OY 21 SerAlaProValThrIysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgLVal 40
   ::::::::::
Db 1420 AATGCCCGAGTTCAAACAGCTCCGAGTACAGCGTGGCCATCATCAGCTGACAGCAGT 1479
OY 41 SerLeuArgMetCysGln---ThrAlaValSerAspLysIleSerIleAspTyrValTyr 59
   ::::::::::
Db 1480 GCGTTTGCCCTTCACCTTCATCCAGGTGGTGACAG----- 1518
OY 60 ArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGlu-----GlySer 75
   ::::::::::
Db 1519 ---GATAGAAATTTGGGCTGAAACAGCATTTTGAGGTGACTTGGTGGGAAC 1569
OY 76 ProAlaTyrPheTyrValAlaAsnArgLysAsnAsnGlnGlyTyrAlaLeuValAlaAla 95
   ::::::::::
Db 1570 AACTCCACCACTTCATCATCTCCCGACCTCCGTCAGGGAGGAGGAGCAGTCTGAT 1629
OY 96 AspAspArgIleProThrIleLeuAlaTyrSerProIleGlyThrPheAspMetAsp--- 114
   ::::::::::
Db 1630 CCGGTGGCCATCCCA-----CTGGACTACGAGACCGTGGACCGCTACGACTTTGATCTC 1683
OY 115 -----SerMetProAspAsnLeu-----ArgMetThrLeuGlnIleTyr 127
   ::::::::::
Db 1684 TTTCGCCAATGAGAGTGTGCTGACATGTGGCTATGCCAAGTGAAGTCACTTCATC 1743
OY 128 AspGlnGluIleGlyLeuIleLeuSerGlyLysAlaGlnLeuAsnGlnGluIleLeuArg 147
   ::::::::::
Db 1744 AATGAAATGACAAACGCGCCACTTCAGCCAGCCACTGTACAAACATCAGCTGTACGAG 1803
OY 148 ThrGlnGlyAlaProAlaGluValHisAlaLeuMet-----AspAsnGly----- 162
Db 1804 AACGTACCGTGGGAGCTCTGTGTGACAGCTCTGGCAACATGACATGATGACGAGCACC 1863
OY 163 -----HisPheAlaAsnAspProMetArgTrpAsnGlnGlyTyrPro 176
   ::::::::::
Db 1864 TTTCGGGAACTACAGTACTTCTTCAGTATGATACCTGACAGCTTCTGCTG----- 1914
OY 177 TrpAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAla 196
   ::::::::::
Db 1915 ---GACAAAGGAC-----ACGGACTCATCATG 1938
OY 197 ThrAlaAla-----AlaGlnIleMetArgTyrHisSerTrpProLeuGlnGly 212
   ::::::::::
Db 1939 CTGATTCGCGAGCTGACTATGAGCTATGCTCAGCGCTCAGCCCTGACGATGATCCCGG 1998
OY 213 GlnGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGly 232
   ::::::::::
Db 1999 GACGGGGCGCGGACGAGACCAAGCGCGGTGACGATCAAT----- 2040
OY 233 GluMetTyrAspTrpIleAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSer 252
   ::::::::::
Db 2041 ---GTGTGGAT---GTCAAC-----GACAAAGGCGCCACTTC 2073
OY 253 GlnValAspAlaLeuTyrAlaThrIleuMetArgAspValSerAlaSerMet----- 270
   ::::::::::
Db 2074 CAAAGAGATGCTTACCTGGGTGCTGCGGAGAACGAGCTTCTCAGACAGCTGAG 2133
OY 271 -----SerPheTyrGluAsnGlySerGlyThrTyrSerValTyr 283
   ::::::::::
Db 2134 CCGCTCCGGGCAACAGATGAAGACTCCCTCCCAACACCAAGTCACTCAAGCATTTGTC 2193
OY 284 ValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeuHisValArg 303
   ::::::::::
Db 2194 AGTCATCTGCTTGGACGCTACTTCGACATC----- 2226
OY 304 AlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyGluLeuAlaSerGlyArg 323
   ::::::::::
Db 2227 AGCTGTACGAGGCTAT-----GGAGTGTACAGCGCTGACGCTGCG 2265
OY 324 ProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyr 343
   ::::::::::
Db 2266 CCCCTGATATTACAA----- 2280

```

```

OY 344 AlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyValSerAsnGlyPheTyr 363
   ::::::::::
Db 2281 -----CAGATATCAATGGCGTAT 2301
OY 364 LysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyValLeuGlyIleGlyPheThr 383
   ::::::::::
Db 2302 TATCTACGAGTCAATG-----GCCATGATGATGCGGACACCCCTCTCAACAGCAC 2352
OY 384 IleTyrGlnGluIleThrIleThrGlyIleGluProAlaLeuTyrProAla----- 399
   ::::::::::
Db 2353 GTCCCTGTACCATTCAGAGGTGTGATGAAATGACAAACCTCCACCTTCAGCAAGCCC 2412
OY 400 -----GluAlaGlyThrAspAlaLeuProIleLeu 409
   ::::::::::
Db 2413 GCCTACTGTCTCCCTGTGGAGAACATCATGTCAGAGACCAAGCTGCTGTCTGAT 2472
OY 410 AlaLeuLys---AspIleGluAlaGluTyrLysSerGlnSerGlyLeuAsnAlaGlyTyr 428
   ::::::::::
Db 2473 GCCACAGACCTGACCGCTCCCGGAGTACGGCCAGAGTCC-----ATCATCTAC 2523
OY 429 SerIleTyrAsnThrGlyGluGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLys 448
   ::::::::::
Db 2524 TCCTTG-----GAAAGCTCCACCCAG-----TTTGGATCATATGCC 2559
OY 449 AlaAspGlyGluValIle-----GluValLysThrSerIle 461
   ::::::::::
Db 2560 CCGTCAAGGGAATCACCACACAGCTCTGTGACCGAGAGACCAAGCTGTGATCATAC 2619
OY 462 AsnIleSerTrpTyr-----GlyTyrGlnGlnHisProGlnSerPheSerLeuAla 478
   ::::::::::
Db 2620 CTCTATGTTGCGGCACTGACGGGTGTGGGACACACAGAAACT----- 2667
OY 479 ProAsnGlnLeuSerGlnGlyIleAsnThr-----IleThrLeuLeuTyrArgArgThr 496
   ::::::::::
Db 2668 -----GECATGCGCCACCGTAAATCATCACTCTCTGACATCATATGAC 2709
OY 497 GlyThrGlnGlnTrpGluProValArgHisAlaGlnGlyGlyTyrValAsnSerIleLys 516
   ::::::::::
Db 2710 AACCAACCCACGTCGAGACGACGCCCTTAC-----TACATCAACCTGCTGGAG 2757
OY 517 ValAsnThrThrAspProAsn---AsnValValValThrValAspAsnAsnGlnGlyLys 535
   ::::::::::
Db 2758 ATGACCCCTTCACACCTGATGTGACCCAGCGGTGTGGTGTGACCAACCTGCGGGAG 2817
OY 536 -----LeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyr 550
   ::::::::::
Db 2818 AATGGACACCTGCTGACAGCATCCAGCCACCAACAAATTCACAGCTCAACAGCAC 2877
OY 551 -----GluHisSerThrIleThrValGlnPheAsnSerAspSerProAsp----- 565
   ::::::::::
Db 2878 ACGGCAAGATCCGACACACCCAGCCATGCTGGAGCGGAGAACCCGACCCCATGAG 2937
OY 566 -----GluIleArgThrProValAlaPhe 573
   ::::::::::
Db 2938 GCCGAGCTAGCGCAAAATCGTCTCTGTTACTGACTGGGACAGCCCTCTGAAA 2997
OY 574 AlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeu----- 587
   ::::::::::
Db 2998 GCCACAGAGTCCACACATGTTTGAAACCTTGGATCTGAATGACAAATGACCCACC 3057
OY 588 -----GlyTrpValMetAlaGluValProGlyLysSerSerAsn 600
   ::::::::::
Db 3058 TTTCAGAACCTGCTTGTGCGCGAGGCTTGAAGGATCCCGCGGGGTCTTCATC 3117
OY 601 TyrProValValTrpSerLysAspValLeuThrLeuSerGlnGlyAspTyrThrLeuTrp 620
   ::::::::::
Db 3118 TACCAAGTGTG-----GCCATGCACTGTGATGAGG----- 3150
OY 621 TyrArgPheSerIleAsnAsnGlnLysAspGluTrpLysIleGlySerValSerVal 640
   ::::::::::
Db 3151 -----CTGAC-----GCGCTGTGTCTCTAC 3171
OY 641 LysThrProThrGluTyrThrHisProLeuPheGluValGlyHisAsnGlnThrSerThr 660

```

```

DB 3172 CGCATCGCGGGGCGCATGCGCCGCGATGCGCTTCGATCAACAGACGCGCGCGTGC 3231
OY 661 TTTTLeuaspMetAlaHisnArgValLeuProaspPheThrLeuys----- 677
DB 3232 GTCAACACCGACCGAGCTGCGCGGCGATCGCGGATACCGACTCGCGGTGGTGC 3291
OY 678 ---AsnLeuGlyLeuProPheasnGlyGluLeuValValPheArgGlnThrGlnSer 696
DB 3292 AGTGTGCGAGCGCGCC----- 3318
OY 697 SerSerGlySerLeuTrpAlaAlaGlnGluThrValHisIle-----LysGln 712
DB 3319 TCCACCGACGCGCTC-----ACCATCCATGCTCTGATGTGACAGCAGG 3363
OY 713 GlyLeuThrPhe-----ValTrpLysProValValGlnGlyProIlePro----- 727
DB 3364 ACGCCACCTTCTTCCCGCGGTGACATGTCTGTGTCGAGAGCGCGCGCGAG 3423
OY 728 -----AspGlySerTrpArgAlaThrLeuHis 736
DB 3424 TTCGGGCTGCTCTGCGTGAAGTGCAGGACGACGAGCGCGCTCAATGCGAGCTCAGC 3483
OY 737 AlaPheValAsnGly-----GlnGlnGlnLeuTrpLeuGlyLysArgAsnTrpThr 754
DB 3484 TACTTCATCAGACAGGTGGCAAGTGGAGTGGAAATTCACAGCGTGGTTCACGCGCTT 3543
OY 755 ValLysIleValAsnGly-----ThrAla-----ValGlu 764
DB 3544 GTGAGAACCGTGTGGCGCTGCGAGCGGAGACGACGCGCGCTCATGCTCATCTCGAG 3603
OY 765 AlaIleGlySerSerGlnGluIleArgValPheProAsnProAlaArgAspTrpVal--- 783
DB 3604 GCCATCGACAGAGCGCCCTGTAGGAGAGACACAGCGGACAGCCAGCTGTTCGACT 3663
OY 784 -----GluIleSerAla 787
DB 3664 GTCTGTGATGTAATGACACCGGCCCATCTTCTGACAGACGAGTATGAGCGCGCTC 3723
OY 788 Pro---CysIleProGlnGluThrSerIleIleLeuPheAspLeuSer----- 802
DB 3724 CCTGAGGACATCCCTGAGCGCACGACTCTTGCAGCTGMAAGCCAGCCAGCATGAG 3783
OY 803 -----GlyLysIleValMetLysAsnSerLeuSerAlaGlyHisGly----- 816
DB 3784 GCGGAGTTGGGCGG---GTGTGTACCGCATCTTCATGTACCATGCGCAACAACCTTC 3840
OY 817 ArgMetAspValSer-----ArgLeuPro 824
DB 3841 CGGATCCATGTACGCAATGGGCTCTGATGCGAGGGCGCGCGCTGAGCGGAGCGG 3900
OY 825 AsnGlyAlaTrpIleLeuLysValAspGlyTrp 835
DB 3901 AACTCATCCAGCTGCTGATGTAGTGAGCGCTAC 3933

```

```

MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 489:
SEQUENCE CHARACTERISTICS:
LENGTH: 7712 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-070-927A-489
SEQUENCE DESCRIPTION: SEQ ID NO: 489:
Alignment Scores:
Pred. No.: 0.00367 Length: 7712
Score: 133.50 Matches: 143
Percent Similarity: 33.29% Conservative: 94
Best Local Similarity: 20.08% Mismatches: 272
Query Match: 3.01% Indels: 205
Gaps: 34
US-10-030-330-1 (1-843) x US-09-070-927A-489 (1-7712)
OY 131 ILeGlyLeuIleLeuSerGlyLysAlaGlnLeuAsnGlnGluLeuArgThrGlnGly 150
DB 6820 ATTAGGCTTAGCTTGGAGGCGACAGAAAAGAACCGCGCTCGGAATTCAGACAA 6761
OY 151 ValProAlaGluValHisAlaLeuMetAspAsnGlyHisPheAlaAsnAspPrometArg 170
DB 6760 GTTAAAGGAGAGCTACATATTATGATGAAGAAAAGAAATGATGCGCAAGT----- 6713
OY 171 TPAsnGlnGlyTrpProTrpAsnAsnLysGluProLeuLeuProAsnGlnLysAla 190
DB 6712 -----TTATTAAGTTGGTACGCGAGTC 6692
OY 191 TTTThGlyCysValAlaThrAlaAlaGlnIleMetArgTrpHisSerTrpProLeu 210
DB 6691 GTTGCCGCTAGTTTACACACACTTTCGGCAGCGCTGTAACGACAGGAATACCGCAGTG 6632
OY 211 GlnGlyGlnGlySerPheAspTrpHisAlaGlySerLeuValGlnAsnTrpSerGlyThr 230
DB 6631 CAGGTGGAG-----TTGTGTGGGAGAACATTAATACGACGACGAGGAAT 6587
OY 231 PheGlyGluMetTrpAspTrpIleAsnMetProGlyAsnProAspLeuAsnLeuThr 250
DB 6586 GCTGGAGAT-----AACACAGCACCGACCGCGCAGTACCTTACCTAAT 6542
OY 251 GlnSerGlnValAsp---AlaTrpAlaThrLeuMetArgAspValSerAlaSerValSer 269
DB 6541 TTCGACTTAGTAAGTACCAAGAGAGTATTAACACCAACCAAAATTAAGTATAGT 6482
OY 270 MetSerPheTrpGlnAsnGlySerGlyThr-----TyrSerValTrpValValGlyAla 287
DB 6481 CTA-----CAAAATATTAGTTCTACACAAACAATATCTGCTATGCAATCGAGAT 6431

```

RESULT 18
 US-09-070-927A-489/C
 Sequence 489, Application US/09070927A
 Patent No. US20020120116A1
 GENERAL INFORMATION:
 APPLICANT: Charles A. Kunsch
 Patrick J. Dillon
 Steven Barash
 TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 NUMBER OF SEQUENCES: 982
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:

```

OY 288 LeuArgAsnAsnPhenArgTyrLysArgSerLeuGlnLeuHisValAlaGlnLeuTyrThr 307
    :|||
    :|||
Db 6430 ATAGAGGA-----ACG 6419
OY 308 SerGlnGluThrHisAspMetIleArgGlyGlnLeuAlaSerGlyArgProValIlyTyr 327
    :|||
    :|||
Db 6418 AAGAAAGGGGTGACAGCTTCAGCAAACTTGGCTAATGGAGTGAAGCC---TTA 6362
OY 328 AlaGly-----AsnAsnGlnSerIleGlyHisAlaPheValLysAspGly----- 342
    :|||
    :|||
Db 6361 GCTGGAAACCTTAATTTAGTCAAAATCGGATATTATTCAAAATTGATGCAACAAAACA 6302
OY 343 -----TyrAlaSerAspGlyThrPheHisPheAsnThrPglYThrPglYValSer 359
    :|||
    :|||
Db 6301 ATATATGTTTATTTAATCTGATCTGAGAGTTTACGTATGTAATGACCTGCAATGCA 6242
OY 360 AsnGlyPheTyrLysLeuThrLeu-----LeuSerProThrSerLeuGlyIleGly----- 376
    :|||
    :|||
Db 6241 CCTGATTTTTCGGGAACGAATATTCGCTTAAGCGGTGATGCTATTCGAATTCGACAGCA 6182
OY 377 -----GlyGlnGlyIleGlyPhe-----ThrIleTyrGlnGlnIle 388
    :|||
    :|||
Db 6181 AGTTCAAGAAAGACAGACAGCATATGTATTCAGTATTCGATTCATGATCTTGAAGTT 6122
OY 389 IleThrGlyIleGluProAlaLysThrProAlaGlnAlaGlyThrAspAlaLeuProIle 408
    :|||
    :|||
Db 6121 ACTACCTCCATACCAAAATTCAAAGCAGAGTCCCTACACAGCAACATCATGCAACCTC 6062
OY 409 LeuAlaLeuLysAspIleGlnAlaGluTyrLysSerGlnSerGlyLeuAsnValGlyTyr 428
    :|||
    :|||
Db 6061 GTCGCAAGACCATCAATTTAAACACAGC-AMCGAAACAGAGAGGAGAAACCTCACCTAT 6003
OY 429 SerIleTyr-----AsnThrGlnGlnGlnGlnSer 438
    :|||
    :|||
Db 6002 TCTGTCTCT-CTAATAATCATTCATTCGTAAGAGAGAAAGCAATGAAAAAGAAAT 5544
OY 439 AsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluValIleGluValLysThr 458
    :|||
    :|||
Db 5943 -----GATGGCAAGTT----- 5932
OY 459 SerSerIleAsnIleSerTrrTrrGlyTyrGlyGlnHisProGlnSerPheSerLeuAla 478
    :|||
    :|||
Db 5931 -----ATTAGTTGGTAGCCAGCTGTAAGCCGCAAGTTTATGACCTTTTACAG 5884
OY 479 ProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThr 498
    :|||
    :|||
Db 5883 -----ACAAACCGGT 5875
OY 499 GlnGlnTrrPglProValArgHis-AlaGln-----GlyGlyTyrVal-----As 513
    :|||
    :|||
Db 5874 GACCACTGGGAATACACACAGTCCACAGCAATTTGCAGCGCGCAGCTGCCAGATGGTAA 5815
OY 513 nSerIleLysValAsnThrThrAspProAsnAsnValValValThrValAspAsnGln 533
    :|||
    :|||
Db 5814 TGGCGTTCCA---AATACAGCTGCATCCAGAC-----CCGACCAAGCCAAATAGTAA 5767
OY 533 uGlyLysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSe 553
    :|||
    :|||
Db 5766 CTTTGATTTACTTTTATTTCA-----AGCAATATCATTTTGG 5728
OY 553 rThrIleThrValGlnPheAsnSerAspSerProAspGlnIleArgThrProValAlaPhe 573
    :|||
    :|||
Db 5727 GAATATATCTATTTCT-----GATGATTTTAACTAAACCAATTCCTTAA 5686
OY 573 eAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTrrValMetAlaGln 593
    :|||
    :|||
Db 5685 TAAACAGATGAAGGA---CGTGTCTGGAATGTGGAATCTGTGGT-----GTAGAGAGA 5635
OY 593 uValProGlyLysSerSerAsnThrProValIleThrPserLysAspValLeuThrLeuSe 613
    :|||
    :|||
Db 5634 TATTCAGGAGGAGCAAAAGAGCTGCGATGCTGCTCAATGCACTGGTATGAAGCTCGG 5575
OY 613 rGlnGlyAspTyrThrLeuThrPtyrArgPheSerIleAsnAsnGlnLysAspGluThrPly 633

```

```

Db 5574 CGAGGAAAGTTTG----- 5562
OY 633 slYsIleGlySerValSerValLysThrProThrGluTyrThrHisProLeuPheGluVa 653
    :|||
    :|||
Db 5561 -----GAAGCAGATTTTAAACAGAGAGCAATGCTTTTGTATTCATTA-----AT 5518
OY 653 lGlnHisAsnGlnThrSerThrTyr-----ThrLeuAspMetAlaHisAsnArgVa 670
    :|||
    :|||
Db 5517 TTTATATGAGGAACATAAACGATATGATATTCAAATGCGATCTGCGGACCAAGTTAC 5458
OY 670 lleuPro-----AspPheThrLeuLysAsnLeuGlyLeuProPheAsnGlyGlu 687
    :|||
    :|||
Db 5457 CAAAACGAAACTGTGTGATACAAACACCTTTTCAATGCTTTA-----GGCGAGATGC 5404
OY 687 uValValValPheArgGlnThrGlnSerSerSerGlySerLeuThrAlaAlaGlnGlnThr 707
    :|||
    :|||
Db 5403 AGTGTCTTTCGAAATGCTACACAGCAAAAGGCGCAGAGATTTGG----- 5358
OY 707 rValHisIleLysGlnGlyGluThrPheValTyrLysProValValGlnGlyProIleP 727
    :|||
    :|||
Db 5357 -----CAATTTTCATTTTACCATATTTTAAATTAAGCAAC 5320
OY 727 oAspGlySerTyrArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnLeuTyrLe 747
    :|||
    :|||
Db 5319 A-----GCTTCAATATCAAAAGCAGAGTCCCTACACAGC----- 5286
OY 747 uLysGlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaValAlaIleGln 767
    :|||
    :|||
Db 5285 -----AATATCATCTTGGAACTTACGCGAGGACCATCAATTTTAAAGCAAC 5239
OY 767 uSerSerGlnGluIleArgValPheProAsnProAlaAspTyrValGlnIleSerAl 787
    :|||
    :|||
Db 5238 AACAAAGAAAGAC-----AGACAGGACCATCAGCGATTAAGTGC 5200
OY 787 aProCysIleProGlnGlnThrSerIleLeu 798
    :|||
    :|||
Db 5199 TATGTGCTGCTCCCTAATATCATTCATTCGTA 5166

RESULT 19
US-09-974-300-2632
; Sequence 2632, Application US/09974300
; Patent No. US2002014672A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974, 300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680, 598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279, 526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2632
; LENGTH: 4557
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2632

Alignment Scores:
Score: 0.00186 Length: 4557
Percent Similarity: 133.00 Matches: 193
Best Local Similarity: 32.17% Conservative: 129
Query Match: 19.28% Mismatches: 351
DB: 3.00% Indels: 329
Gaps: 51
US-10-030-330-1 (1-843) x US-09-974-300-2632 (1-4557)

```


DB 1921 ACTGAACGCAAAATAG-----GAAGCAGCAGTTTACGCCGATCCG 1965
QY 646 TyrThr-----HisProLeuphegluValGlyHisAsnGlnThrSerThrTyrThrLeuAsp 664
DB 1966 CAGACATCTCCACACAGGCTTCAACACGCG----- 1995
QY 665 MetAlaHisAsnArgValLeuProAspPheThrLeuValAsnLeuGlyLeuProPheAsn 684
DB 1996 ---TTATACACAGCTGTTCAAAAGAAATCAGGTGAAA---ATGGCATCAACTATAT 2049
QY 685 GlyGlu----- 686
DB 2050 GGAGGCGCTCAAAAACCTATATTAAGATTCTTACAGATGACGACCAATTGAA 2109
QY 687 -----LeuValValPheArgGlnThrGlnSerSerGlySerLeuTrp----- 702
DB 2110 AAGGCTCGTGTGCTCAAGACCTATTCAGTTAATAAGACGGTTCATCAACAGAGA 2169
QY 703 -----AlaAlaGlnGlu----- 706
DB 2170 GACATCTGCGCGCTCCCAATATGATGTGGAAGCGCTGCGCTCCCAATAACAAC 2229
QY 707 ---ThrValHisLeuGlnGlnGlyLutThrPheValTyr-----LysProVal 721
DB 2230 TTGACCGCTCATTTGAAAAGCAGCAGATTCCTCCGCTATTTGAAATTTAAACATCG 2289
QY 722 ValGlnGlyProIle-----ProAspGlySerTyrArgAlaThrLeuHis----- 736
DB 2290 CTAAAGACAGCTATTAGACACGATCTTACACCAATAGGCCACGTAACATTAATGCC 2349
QY 737 -----AlaPheValAsnGlyGlnGlnLeu 745
DB 2350 GATATTCAGACGCTGAGTTGACGCGATCTTTCGCTGCGACAGCAGGAAAGCTCGTT 2409
QY 746 TyrLeuValGlyLysArgAsn-----TyrThrValLysIleValAsnGlyThrAlaVal 763
DB 2410 TTCAAGAGGACCAAAACGAGGCTTGTGCGATTGAGCAGCATCAATGTCACACGCGAC 2469
QY 764 GluAlaIleGlnSerSerGlnGluIleArgValPheProAsnProAlaArgAspTyrVal 783
DB 2470 CAGTCGGTG-----TTGGAAGATGTAAAGTCACGATACCTCCG----- 2508
QY 784 GluIleSerAlaProCysIleProGlnGluThrSerIleIleLeuPheAsnLeuSerGly 803
DB 2509 -----GATACCAATCAAAATTTTACCTGAAGATCTTTT 2541
QY 804 LysIleValMet-----LysAsnSerLeuSerAlaGlyHisGlyArgMet 818
DB 2542 AAGATTATTCAGCCTTAATATGATGAAAAGGTGCGTCAAAAGACAGCAGCGAAATCTT 2601
QY 819 AspValSerArgLeuProAsnGlyAlaTyrIleLeuLysValAspGlyTyrThrThrLys 838
DB 2602 -----GTCCCGGATGATGTTCAGCTTAAGAAAAGGAGACATATACACGTGAT 2649
QY 839 Ile 839
DB 2650 ATC 2652

RESULT 20
US-10-114-170-206/C
Sequence 206, Application US/10114170
Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Butland, Valerie
Berne, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. US20030023075A1 Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSER: Quarles & Brady
STREET: 1 South Plinckney Street

CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-Dec-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-Dec-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 206:
SEQUENCE CHARACTERISTICS:
LENGTH: 43360
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 206:
US-10-114-170-206
Alignment Scores:
Pred. No.: 0.0578 Length: 43360
Score: 133.00 Matches: 165
Percent Similarity: 30.62% Conservative: 128
Best Local Similarity: 17.24% Mismatches: 338
Query Match: 3.00% Indels: 326
Gaps: 44
DB: 9
US-10-030-330-1 (1-843) x US-10-114-170-206 (1-43360)
QY 74 GlySerProAlaTyrPheTyrValAlaAsnArg-----GlyAsnAsnGlnGlyTyrAla 91
DB 18309 GGTTAGCATTTGTTTATGTATTAATTCAGAAATTCATTGCAAAAGTGCAAAATATTTT 18250
QY 92 LeuValAlaAlaAspArgIleProThrIleLeuAlaTyrSerProIleGlyArgPhe 111
DB 18249 AATATGGTTCGATTCAAAACGTAACTGATGATAGCTATCAG-----AATGCCCTT 18196
QY 112 -----AspMetAspSerMetProAsp-----Asn 119
DB 18195 TTTATACCTGAAAACGTGTGAACCTGTGGCATCTTTCRAATCCGCAAGATATTTAT 18136
QY 120 Leu-----ArgMetTrp-----LeuGlnIleTyrAspGlnGluIleGlyLeuIleLeu 135
DB 18135 TTATGACAGCATTTGCTGTTGAATAAGCATTTATACAGTTCTGAAACCAATGATGAAG 18076
QY 136 SerGlyLysAlaGlnLeuAsnGlnGluIleLeuAsnGlyThrGlnGlyValProAlaGluVal 155
DB 18075 GCCGGCGCTGCGTACAG-----CAGATCATTTTCCACACGCAAAAACCTTCTTGAAATAC 18022
QY 156 HisAlaLeuMetAspAsnGlyHisPheAlaAsnAspPrometaTyrTrpAsnGlnGlyTyr 175
DB 18021 ACTGCACATA----- 18013
QY 176 ProTyrAsnAsnLysGluProLeuProAsnGly----- 187
DB 18012 CCACCTTTAAGTTGCGACACTCTTGTGCTGACAGGTGCTTGTGCTGACACAGAAATAA 17953

OY 188 -----AsnHisIaIaTyrThr 192
 Db 17952 CTGACTAAATATGTCCTCCGAGCATGACCAACAAACATGACCATGACCAAGCATTAAT 17893
 OY 193 GlyCysValAlaIleThAlaAlaIleMetArgTyrHisSerTrpProLeuGlnGly 212
 Db 17892 TATGGCGGCACAAACAGCGCGAGTCTCGTAGCCAGCTTCAGCGAGTCTGACAGCGC 17833
 OY 213 GluGlySerPheAspArgTyrHisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGly 232
 Db 17832 GATTACGGCAAGATACCGCTCTTGGT---ATCGCTGTACCGAGCTTCGTA----- 17782
 OY 233 GluMetTyrAspTrpIleAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSer 252
 Db 17781 CAGTTGCAGCGCGCTGTACACATTAATGACCGGAGGTAAATGTCGACAGTGTAAAT 17722
 OY 253 GluValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPhe 272
 Db 17721 AACTTTCAGCGT-----ACTTCACTGCACTTCTTATTAACCGCTTC 17683
 OY 273 TyrGluAsn----- 275
 Db 17682 TATGATTCGCAAAATATGTCGATTTGGTCTGAGTGCAGCGGCTTACATTCAGTCCCGC 17623
 OY 276 -----GlySerGlyThr-----Tyr 280
 Db 17622 TTACCGCAAAATTTAGGTGCGGGTACGCTTTTCTCTCTGCAACAAATGTTGGCTAAT 17563
 OY 281 SerValTyrVal-----ValGlyAlaLeu 288
 Db 17562 AACGCTTCATGATGATGATGATTTCTGTGATATACCGCTTACGATTAATGTCGCGCA 17503
 OY 289 ArgAsnAsnPheAsnGlyTyrLeuAspSerLeuGlnLeuHisValAlaArgAlaLeuTyrThrSer 308
 Db 17502 TACTGCGCGAGCATTTTCAAAAGTAC-----GTAAACCGCTATTTTCCGCAATG 17455
 OY 309 GlnGluTrpHisAspMetIleArgGlyGluLeuAlaSerGlyArgProValTyr----- 326
 Db 17454 ACCGGCTGCGATGAGCATACAAATAAGAAAGCATATGATGACGCGCCACCAAAATGGCTTC 17395
 OY 327 -----TyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCys 340
 Db 17394 GATATCCGTTTAAATGGCTATATACCGCTATATCCGCAATGAGCCCAACGATATATAT 17335
 OY 341 AspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrPrlGlyAlaSerAsn 360
 Db 17334 GAGCGATATATATGATATATGCTTGTATTAATTCGAT----- 17293
 OY 361 GlyPheTyrLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlnGlnGlyIle 380
 Db 17292 -----AAGCTG-----CAGTCGAATCCGTGCGCGCAGCCGCTGGTGTAT 17254
 OY 381 GlyPheThrIleTyrGlnGlnIleIleThrGlyIleGlnProAlaLysThrProAlaGln 400
 Db 17253 AACTATACATCCGATTCCTCTGCTGAGATGGGATC----- 17218
 OY 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlnAlaGlnTyrLysSer 420
 Db 17217 -----GATTTCGCT----- 17209
 OY 421 GluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlnGlnGlnSer----- 438
 Db 17208 -----CATGTCACGCGTAAATGATATATCTCTCT 17179
 OY 439 ---AsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlnValIleGlnValLys 457
 Db 17178 TACTCAATGCGACTTCGCTATGATGATTAATGCTGCTGACGAAATTAAGACACAG 17119
 OY 458 ThrSerSerIleAsnIleSerTrpTyrGlyTyrGlnHisProGluSerPheSerLeu 477
 Db 17118 -----TATGTTAACGACTTAAGAAACATTAATGAGC 17089
 OY 478 AlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgArg----- 495

Db 17088 AGCGCTGATGATGCTGATGATGATTAATGATTAATGATGATGATTAATGATTAATGAT 17029
 OY 496 -----ThrGlyThrGlnGln----- 500
 Db 17028 ATTCTTCTCTGATATATTCGATGATATATTAATGATTAATGATTAATGATTAATGAT 16969
 OY 501 -----Trp---GluProValArg 505
 Db 16968 CAGTTATGCTTAAGCAAAATACGCTGATGATGATGATGATGATGATGATGATGATGAT 16909
 OY 506 HisAlaGlnGlyGlyTyrValAlaAsnSerIleLysValAsnThrThrAspProAsnVal 525
 Db 16908 CGCAGTCAGCGCGCTGATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 16849
 OY 526 ValValThrValAlaAsnAsnGlnGlyLysLeuSerIleValAlaProAsnSerPheValAla 545
 Db 16848 ATTTTCTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 16804
 OY 546 AspLeuAsnSerTyrGlnHisSer-----ThrIleThr 556
 Db 16803 AGCGCTGCGCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 16744
 OY 557 ValGlnPheAsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSer 576
 Db 16743 GTTCTGTCGAATGTCGAATGTCGAATGTCGAATGTCGAATGTCGAATGTCGAATGTCGAAT 16684
 OY 577 ThrGlyAlaThrAlaAspAspValIleSerLeuGlyTyrValMetAlaGlnValProGly 596
 Db 16683 ACTTCGCTTAACGATTAACGCGGATACCATTAATGATTAATGATTAATGATTAATGAT 16624
 OY 597 Gly-----SerSerAsnTyrProValValIleTrpSerLysAspValLeuThrLeuSerGln 614
 Db 16623 GGGGTACGTCAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 16579
 OY 615 GlyAspTyrThrLeuTyrTyrArgPheSerIleAsnAsnGlnLysAspGluTyrLysLys 634
 Db 16578 GGAACGCAACGCTT-----GGGCAAAATATGTCGCAAAACGATGCAACGCT 16531
 OY 635 IleGlySerValSerValLysThrProThrGlnTyrThrHisProLeuPheGluValGly 654
 Db 16530 AAGCAACGCTTAACGCTTAACGCTTAACGCTTAACGCTTAACGCTTAACGCTTAACGCT 16495
 OY 655 HisAsnGlnThrSerThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAspPhe 674
 Db 16494 CAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 16456
 OY 675 ThrLeuLysAsnLeuGlyLeuProPheAsnGlyGluValValAlaPheArgGlnThr 694
 Db 16455 -----GCACCTTAATGCCAGTCCGCTTATATTTTGTATCAAAAC 16417
 OY 695 GlnSerSerSerGlySerLeuThrPheAlaGlnGlnThrValHisIleLysGlnGlyGln 714
 Db 16416 AAGCGCAGCATTTCTAGAT---AAGCGCAGATTAAGCAACGTCAGTCAAGTAAATGCTAAG 16360
 OY 715 -----ThrPheValTyrLysProValValGlnGly---ProIleProAspGlySer 730
 Db 16359 GATGCTATTAATATATCTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 16300
 OY 731 TyrArg---AlaThrLeuHisAlaPheValAsnGlyGlnGlnIleLeuTyrLeu----- 747
 Db 16299 GTTACATTTTCACAAACATTTGGGATGTCACAGGTAAGTCGCAACGCAACGCAACGCA 16240
 OY 748 -----LysGlyLysArgAsnTyr 753
 Db 16239 GGAATGATGCTGCTGCGAGATTAACGATTAACGATTAACGATTAACGATTAACGATTAAC 16180
 OY 754 ThrValLysIleValAsnGlnTyrThrAlaValGlnAlaIleGluSer-----GlnGlu 771
 Db 16179 AGTGCACAGTCAGTATGAGGCTGAGGATTAAGCAACGATGAGTCACTTTTGTATGAGAA 16120
 OY 772 IleArgVal-----PheProAsn 777

```

Db      16119 CTGAATATTGACAAACAGGTGATATATTGTAACAATGTCAGAGCGAGTGCCTAAT 16060
OY      . 778 ProAlaArgAspTyrValGluIleSerAlaProCys----- 789
Db      16059 ATTGGCTGCATATGGTCAGTTTAACGAAAGCAAGCGGTGATGTCATATTCA 16000
OY      790 ---IleProGlnGluThrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLys 808
Db      15999 TGGTATTCAGAAAATACCACTATCGGACTGTCGATGCATCAGGGAAAGTCACTTTGAAT 15940
OY      809 AsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeuProAsnGly----- 826
Db      15939 -----GTAAGGCACTGTC---GTAATTAAGCCACATCTGGTGAATAAG 15898
OY      827 -----AlaTyrIleLeuLysValAspGlyTyrThrLysIleAsn 840
          ::::: ::::: ::::: ::::: ::::: :::::
          15897 CAACACAGTAAGTACACTATAAAGCACCGCTGCTATATGATATAAAGTGAT 15847

```

Job completed: June 2, 2003, 01:14:35
 Job time : 676 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - nucleic search, using frame_plus.p2n model

Run on: June 1, 2003, 22:41:13 ; Search time 100 Seconds
(without alignments)
2585.285 Million cell updates/sec

Title: US-10-030-330-1

Perfect score: 4438
Sequence: 1 MKSFLAIVMLFGIAMQGH.....PNGAVILKVDYTTKINIVH 843

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus.p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO/spool/US10030330/runat_23052003_181645_6806/app.query.fasta_1.1031
-DB=Issued_Patents_NA -QEXT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=20
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030330.ecgn.1.1.440.etrnat.23052003_181645_6806 -ICPU=3
-NO_XLPRX -NO_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONCLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA: *
1: /cgn2.6/ptodata/1/lna/5A.COMB.seq: *
2: /cgn2.6/ptodata/1/lna/5B.COMB.seq: *
3: /cgn2.6/ptodata/1/lna/6A.COMB.seq: *
4: /cgn2.6/ptodata/1/lna/6B.COMB.seq: *
5: /cgn2.6/ptodata/1/lna/PCTUS.COMB.seq: *
6: /cgn2.6/ptodata/1/lna/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3624	81.7	8439	4	US-08-221-017B-473 Sequence 473, App
2	484	10.9	1197	5	PCT-US95-11723-7 Sequence 7, Appl
3	484	10.9	1197	5	PCT-US95-11723-7 Sequence 7, Appl
4	484	10.9	1197	5	PCT-US95-11723-7 Sequence 3, Appl
5	417	9.4	566	4	US-09-221-017B-423 Sequence 423, App
6	192.5	4.3	504	4	US-09-221-017B-129 Sequence 129, Appl
7	158	3.6	656	4	US-09-221-017B-37 Sequence 37, Appl
8	145	3.3	25165	4	US-08-453-702B-39 Sequence 39, Appl
9	144.5	3.3	4118	1	US-08-119-125A-3 Sequence 3, Appl
10	140	3.2	3106	4	US-08-840-466A-21 Sequence 21, Appl
11	140	3.2	3106	4	US-09-696-188B-21 Sequence 21, Appl
12	140	3.2	9432	1	US-08-277-231A-1 Sequence 1, Appl

13	140	3.2	9432	2	US-08-473-750-4	Sequence 4, Appl
14	140	3.2	9432	2	US-08-477-326-4	Sequence 4, Appl
15	140	3.2	30549	4	US-09-134-001C-322	Sequence 322, App
16	139.5	3.1	2848	2	US-08-805-918-1	Sequence 1, Appl
17	139.5	3.1	3695	4	US-09-453-702B-18	Sequence 18, Appl
18	139.5	3.1	4106	2	US-08-702-572-14	Sequence 14, Appl
19	139.5	3.1	4732	6	5521093-4	Patent No. 5521093
20	138	3.1	4765	1	US-08-750-532-8	Sequence 8, Appl
21	138	3.1	4765	4	US-08-894-818B-7	Sequence 7, Appl
22	138	3.1	4765	4	US-09-445-472-5	Sequence 5, Appl
23	137.5	3.1	4268	4	US-09-453-702B-93	Sequence 93, Appl
24	135	3.0	1666	4	US-09-221-017B-933	Sequence 933, App
25	133	3.0	43360	4	US-09-453-702B-206	Sequence 206, App
26	133	3.0	45325	4	US-09-453-702B-261	Sequence 261, App
27	132.5	3.0	3131	4	US-08-840-466A-20	Sequence 20, Appl
28	132.5	3.0	3131	4	US-09-696-188B-20	Sequence 20, Appl
29	132.5	3.0	4875	1	US-08-460-739-1	Sequence 1, Appl
30	128.5	2.9	3668	4	US-09-206-942-36	Sequence 36, Appl
31	128.5	2.9	3686	4	US-09-206-942-33	Sequence 33, Appl
32	128	2.9	2826	4	US-08-624-655A-1	Sequence 1, Appl
33	128	2.9	15512	2	US-08-853-659A-5	Sequence 5, Appl
34	128	2.9	15512	2	US-08-853-659A-63	Sequence 63, Appl
35	128	2.9	15512	2	US-08-853-659A-66	Sequence 66, Appl
36	128	2.9	24701	2	US-08-853-659A-2	Sequence 2, Appl
37	128	2.9	24701	2	US-08-853-659A-3	Sequence 3, Appl
38	128	2.9	24701	2	US-08-853-659A-60	Sequence 60, Appl
39	128	2.9	24701	2	US-08-853-659A-61	Sequence 61, Appl
40	128	2.9	2037	4	US-08-913-942-14	Sequence 14, Appl
41	127.5	2.9	2079	4	US-09-268-347-25	Sequence 25, Appl
42	127.5	2.9	5629	4	US-09-453-702B-243	Sequence 243, App
43	126.5	2.8	10711	4	US-08-961-527-145	Sequence 145, App
44	126	2.8	8967	2	US-08-853-659A-6	Sequence 6, Appl
45	125.5	2.8				

ALIGNMENTS

RESULT 1
US-09-221-017B-473
Sequence 473, Application US/09221017B
Patent No. 6444799
GENERIC INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
Prior Application DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
Prior Application DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
Prior Application DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
Prior Application DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Monroy, Gladys H
 REGISTRATION NUMBER: 32,430
 REFERENCE/DOCKET NUMBER: 27340-20021.00
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-813-5600
 TELEFAX: 650-494-0792
 TELEX: 706141

INFORMATION FOR SEQ ID NO: 473:

SEQUENCE CHARACTERISTICS:
 LENGTH: 8439 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOHETICAL: NO
 ANTI-SENSE: UNKNOWN
 ORIGINAL SOURCE:

FEATURE:

NAME/KEY: misc_feature
 LOCATION: 1...8439
 US-09-221-017B-473

Alignment Scores:

Pred. No.: 0 Length: 8439
 Score: 3624.00 Matches: 684
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 81.66% Indels: 0
 Gaps: 0

US-10-030-330-1 (1-843) x US-09-221-017B-473 (1-8439)

QY 1 MetLysLysSerPheLeuLeuAlaIleValMetLeuPheGlyIleAlaMetClnGlyHis 20
 Db 6386 ATCAAAAAAGTTTCTTTTACGCAATGCTCTTGGCATTCGCATGAGGAGCAT 6445
 QY 21 SerAlaProValThrLysClnArgAlaLeuSerLeuAlaArgLeuAlaLeuArgLVal 40
 Db 6446 TCGGTCGCGGTACGAAAGAGGAGCTTGAGTCTGGCTCGGCTGGCTTTGCGACAGTA 6505
 QY 41 SerLeuArgMetGlyGlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArg 60
 Db 6506 TCCCTGCGAATGGGCAAAACACAGTATGACAGATTTCCATCGATTACGTTTATCGG 6565
 QY 61 GlnGlyAspAlaGlnArgGlyIleThrSerGlnGlnGlySerProAlaTyrPheTyr 80
 Db 6566 CAAAGGAGTCTGAGAGGGGTATCACATCACAGAGGAGGCTCTCTCGCATATTTTAT 6625
 QY 81 ValAlaAsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIlePro 100
 Db 6626 GTAGCTATCGTGAAATATAGAGGCTATGCTCTTACACACAGATGACAGAAATACCG 6685
 QY 101 ThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAspAsnLeu 120
 Db 6686 ACAATTTTACCTTATACCCATTTGCGCTTGCATGACATGACAGTMTGCCGACATCTT 6745
 QY 121 ArgMetTyrLeuGlnIleTyrAspGlnGlnIleGlyLeuIleLeuSerGlyLysAlaGln 140
 Db 6746 CGCATGTGGCTACAAATTTAGCATCAGAAATAGGCTCTGATCTTCCGAAAGAGCTCAG 6805
 QY 141 LeuAsnGlnGlnIleLeuAlaGlnGlyArgLysAlaProAlaGlnValAlaIleLeuMetAsp 160
 Db 6806 CTCAATGAAAGATATTATACCTACCGAGGCGTACCGCTGAAGTACATCTCTGTATGAT 6865
 QY 161 AsnGlyHisPheAlaAsnAspPheMetArgTyrPheGlnGlnGlyTyrProTyrAsnAsnLys 180
 Db 6866 AACGCTCATTTTCCCAACCATCCCATGCAATGCAAGGTTACCCCATGGAACAATTAAG 6925
 QY 181 GluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAlaIleThrAlaAla 200
 Db 6926 GAACCACTCTCTCTTAATGCAATCATGCTATACGCGCTGTGTTCGCTACTGTCACACA 6985

QY 201 GlnIleMetArgTyrHisSerTyrProLeuGlnGlyGlySerPheAspTyrHisAla 220
 Db 6986 CAATCATCGCGCTACCATAGCTGGCGGCTTCAAGGTGAAGCTCTTTCGATTATCAGTCA 7045
 QY 221 GlySerLeuValGlyAsnTyrSerGlyThrPheGlyGlnMetTyrAspTyrIleAsnMet 240
 Db 7046 GGTCTATTATGTTGGCACTGCTCCGCGACATTTGGTGAAATGTACGACTGATCAATATG 7105
 QY 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAlaAlaTyrAlaThrLeu 260
 Db 7106 CCGGAAATCCCGACCTGATATATCTCAATCTCAAGGATGCTTACGCGCACACTG 7165
 QY 261 MetArgAspValSerAlaSerValSerMetSerPheTyrGlnAsnGlySerGlyThrTyr 280
 Db 7166 ATGCGTATGTGAGGCACTGCTTGTTCATGATGATGATTTTATGAAATGGAAGTGTACTG 7225
 QY 281 SerValTyrValAlaGlyAlaLeuArgAsnAspPheArgTyrLysArgSerLeuGlnLeu 300
 Db 7226 AGCGTTATGTATGATGAGGAGGCTTCCGAAACACTTTCGACAGGCTTCACTGACGCTA 7285
 QY 301 HisValArgAlaLeuTyrThrSerGlnGlnTyrPheAspMetIleArgGlyGlnLeuAla 320
 Db 7286 CATGTACGCGGCTTATATACCTCACAGAGTGGCACGATATGATCCGCGGAGACTTGGC 7345
 QY 321 SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCys 340
 Db 7346 TCCGGAAGGCGGCTATATATGACAGGAATACCCAGCATAGCATAGCTTCTGTTGC 7405
 QY 341 AspGlyTyrAlaSerAspGlyThrPheHisPheAsnTyrPglyTyrGlyValSerAsn 360
 Db 7406 GATGTTATATCTTCGATGATGATCTTCCATTTCACTGGGCTTGGGAGGATTTCCAC 7465
 QY 361 GlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGlyTyr 380
 Db 7466 GGCTTACAACTAACACTCTCCGCGCACTTGGTGGATTCGAGGAGGAGGATATA 7525
 QY 381 GlyPheThrIleTyrGlnGlnIleIleThrGlyIleGlnProAlaLysThrProAlaGln 400
 Db 7526 GGTTTACATTTATCAAGAGATATCAACCGGATTCGAAACCGGCTTACCTCCGCTGAA 7585
 QY 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlnAlaGlyTyrLysSer 420
 Db 7586 GCCGCTACAGAGCTTCCGATCTTGGCATGAAAGCATAGAGCGAGTATTAAGT 7645
 QY 421 GluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlnGlnSerAsnLeu 440
 Db 7646 GAATCCGATTAAGCATAGGCTATTCGATATATATACAGTGAAGAGCAATCAATCTT 7705
 QY 441 AspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlnValIleGlnValLysThrSerSer 460
 Db 7706 GACCTCGATTCAGATTAAGCAAGGCTGACGAGAGTCAATAGAGTGAAGAACTTATCT 7765
 QY 461 IleAsnIleSerTyrTyrGlyTyrGlyGlnHisProGlySerPheSerLeuAlaProAsn 480
 Db 7766 ATCAATATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7825
 QY 481 GluLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgTyrGlyThrGlnGln 500
 Db 7826 CAGTGTCAAGAAAGATCAACACATCACCTTATCGTCGACAGGACCCAGACAG 7885
 QY 501 TyrGlnProValArgHisAlaGlnGlyGlyTyrValAsnSerIleLysValAsnThrThr 520
 Db 7886 TGGAGCGGCTACGCGATGACAGGAGGATATGCAATATAGCATTAAGTAAATACAGCA 7945
 QY 521 AspProAsnAsnValValValThrValAspAsnAsnGlnGlyLysLeuSerIleValPro 540
 Db 7946 GACCCGAAAGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8005
 QY 541 AsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPheAsn 560
 Db 8006 AACAGCTTGTGCGATCTGATTAATCTTATGAAATATGATGATGATGATGATGATGAT 8065

```

Oy      561  SeraspSerProaspGluIleargThrProValAlaPheAlaIleuSerThrGlyAlaThr 580
Db      8066  AGCAGACACCCGTGATGAGATCGTACACCCGTAGCTTTGCTCTATCTACAGAGACTACT 81255
Oy      581  AlaaspaspValIleSerLeuGlyTyrPValMetAlaGluValProGlyGlySerSerAsn 600
Db      8126  GCGGACGATGTAATATCTTTGGCTGGGTATGTGCTGAAGTTCCGGGGGGTAGCAGCAAC 81855
Oy      601  TyrProValValTyrPserLysaspValleuThrIleuSerGluGlyAspTyrThrLeuTyr 620
Db      8186  TATCCGGGGGTTTGCTCTAAAGACGTTCTCCTCTCGGAAGGCGCATATACATTGTGG 8245
Oy      621  TyrTyrPheSerIleAsnAsnGlnLysaspLutryPlysIleGlySerValSerVal 640
Db      8246  TATAGATTTCATCAACACCAACCAAGATGAAATGAAAAAGATCGGAAGCGTGTCACTA 8305
Oy      641  LysThrProThrGluTyrThrHisProLeuPheGluValGlyHisAsnGlnThrSerThr 660
Db      8306  AAAACACCCGACAGATATACGACCCCTTATTCGAAGGGGCGCATATCAACCTTACAC 8365
Oy      661  TyrThrLeuaspMetAlaHisAsnArgValleuProaspPheThrIleuLysAsnLeuGly 680
Db      8366  TATACGCTGGATATGGCACACACAGAGATATGCCCGCACTTTACACTCAAAAATCTCGGA 84255
Oy      681  LeuProPheasn 684
Db      8426  TTGCGCTTTCAT 8437

RESULT 2
US-08-931-320-7
Sequence 7, Application US/08931220
Patent No. 6030635
GENERAL INFORMATION:
APPLICANT: Musser M.D., James M.
APPLICANT: Kapur M.D., Vivek
TITLE OF INVENTION: Methods and Compositions for Identifying
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESS: WEIL, GOTSCHAL & MANGES
STREET: 2882 Sand Hill Road, Suite 280
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025-7022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,220
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/306,542
FILING DATE: 14-SEP-1994
APPLICATION NUMBER: US 08/160,965
FILING DATE: 02-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter Ph.D., Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: BAYL-004/01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6200
TELEFAX: (415) 854-5713
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1197 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

```

	HYPOTHETICAL: NO	
:	ANTI-SENSE: NO	
:	ORIGINAL SOURCE:	
:	ORGANISM: Streptococcus pyogenes	
:	STRAIN: MGAS 1719	
:	IMMEDIATE SOURCE:	
:	CLONE: speB7 (cysteine protease)	
US-08-931-220-7		
 Alignment Scores:		
Pred. No.:	1.7e-41	Length: 1197
Score:	484.00	Matches: 125
Best Local Similarity:	49.62%	Conservative: 72
Percent Similarity:	31.49%	Mismatches: 138
Query Match:	10.91%	Indels: 63
DB:	3	Gaps: 13
 US-10-030-330-1 (1-843) x US-08-931-220-7 (1-1197)		
OY	26 LysGIUAATGAlaLeuSerleuAlaArgLeuAlaLeuArgIuValSerleuArgetMetGly	45
Db	107 AAGAAAGCAAGAATAGCGCTATACACTTTATATCAAAAATCAGCA-GCTATCAAAAGCAGGT	165
OY	46 GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgIngluIyaSpalaglu	65
Db	166 GCACCAAGCGCAGCAAGAT--ATTAAAGCTTGACAAGATTAACTTAGGTGAGAACCTT--	219
OY	66 ArgGLyleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly	85
Db	220 -----TCGGCTCTAATATGTATGTTTACAAATATTCTACTGGA	258
OY	86 AsnaEnglUGlyTYrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr	105
Db	259 -----GGATTGTATCGTTTCAGGAGATAACGTTCCAGAAAATCTTAGATAC	309
OY	106 SerProIleGlyArgPheaspmetaspSerMetProaspasneuarymetTrpLeuGln	125
Db	310 TCTACACAGCGCATTCATTGACCCTTAACGCT--AAAGAAAAACATTCGTTCTTCATGCGAA	366
OY	126 IleTyr-----AspGlnGluIleGlyLeuIleLeuSerGlyLysAla	139
Db	367 AGTTATGTGCAGCAATATCAAAGAAAACAAAATTTAGACACTACTATGCTGGTAACGCT	426
OY	140 GlnLeuasngluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaleumet	159
Db	427 GAGATTAAACAACCAAGCTT-----GTTAATCTCTCTCTT	459
OY	160 AspasnGlyHisPheAlaIsnaspprometArgTrpAsnGlnGlyTyrProTPAsn	179
Db	460 GAT-----TCAAAAGCATTATTCACACCAAGTAACCTTACCAACCTA	504
OY	180 LysGIUProLeuLeuProasn-----GlyAsnHisAla	190
Db	505 TTGACACCGCTGTTATGA AAAAGATAAAACCAAGTGAGCAACTCTTTGAGTCAACATGCA	564
OY	191 TyrThrGlyGcyValAlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTrpProLeu	210
Db	565 GCTACAGAGATGGTGGTCTACTGCCAAGCTGCTCAAAATTTATGAAATATCATATTAATCCCT-	621
OY	211 GlnGIUgLuGlySerPheaspTyr-----	218
Db	622 ---AACAAAGGTTGAAGACACTACACTTGAGACACTAAGTCAAAATACCATATTTCAC	678
OY	219 HisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetYyrAspTrpIle	238
Db	679 CATCCCTAAGAACTTG-----TTTGAGAGTATCTTACTAGACAATACCAACTGAGAC	729
OY	239 Asn---MetProGlyAsnProaspLeuaspasneuleuthGlnSerGlnValaSpalaty	257
Db	730 AACATCTTACTACTTATATGCGGAAGAGATCTAACGTTCAAAAATG-----GGGATT	783
OY	258 AlaThrLeuMetArgspValSerAlaSerValSerMetSerPheTyrGluasnGlySer	277

[illegible]

```

OY 298 leuGlnleuHisValArgAlaLeuGlyThrSerGlnGluThrHisAspMetLeuGly 317
DB 904 GTTCAACCAATTAACCGGACGACCTTACCAACCAAGATTGGACACCAATTAACCAAA 963
OY 318 GluLeuAlaSerGlyAlaArgProValTyrTyrAlaGlyAspAsnGlnSerIleGlyHisAla 337
DB 964 GAATTATGTCACAAACCAACCAACCAATATATCTCAAGGCTGGTGAAGTAGGCGGACATGCC 1023
OY 338 pheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnIleProGlyGly 357
DB 1024 TTTGTATACGATGCTGCTGACGACGACGACGACGACGACGACGACGACGACGACGACG 1083
OY 358 ValSerAsnGlyPheTyrTyrLeuThrLeuLeuSerProThrSerLeuGlyGly 377
DB 1084 GTCTGACGCGCTTCTCGCTGACGACGACGACGACGACGACGACGACGACGACGACGACG 1143
OY 378 GluGlyIleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394
DB 1144 GCGGACGCGCGCTTCAACGCTTACCAAACTGCTGTGTGACGACGACGACGACGACG 1194
OY 1194 GCGGACGCGCGCTTCAACGCTTACCAAACTGCTGTGTGACGACGACGACGACGACG
PCT-US96-05997-3
Sequence 3, Application PC/TUS9605997
GENERAL INFORMATION:
APPLICANT: Musser M.D., James M.
APPLICANT: Kapur M.D., Vivek
APPLICANT: Ananthaswamy, H. N.
APPLICANT: Fernandez, A.
TITLE OF INVENTION: Use of extracellular cysteine protease
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESS: WEILL, GOTSCHAL & MANGERS
STREET: 2882 Sand Hill Road, Suite 280
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025-7022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05997
FILING DATE: 01-MAY-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/306,542
FILING DATE: 14-SEP-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/150,965
FILING DATE: 02-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter Ph.D., Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: BAYL-004/0205
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6200
TELEFAX: (415) 854-3713
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1197 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
STRAIN: MGAS 1719

```

```

: IMMEDIATE SOURCE:
: CLONE: spreb7 (cysteine protease)
: PCT-US96-05997-3
Alignment Scores:
Pred. No.: 1,7e-41 Length: 1197
Score: 484.00 Matches: 125
Percent Similarity: 49.62% Conservative: 72
Best Local Similarity: 31.49% Mismatches: 138
Query Match: 10.91% Indels: 63
DB: 5 Gaps: 13
US-10-030-330-1 (1-843) x PCT-US96-05997-3 (1-1197)
OY 26 LysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnValSerLeuArgMetGly 45
DB 107 AAGACCAAAAGATATACCGCTATCCATTTTCCAAAATATCAGCA-CGTATCAAAAGCAGT 165
OY 46 GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu 65
DB 166 GCACGAGCGCGCAGAGAT--ATTAAAGCTTGACAAAGTTAAGTTCAGGTGGAAGCTT-- 219
OY 66 ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
DB 220 -----TCTGCCTTAATATGATGTTTACATATATTCTACTGCA 258
OY 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
DB 259 -----GCATTTCTTATCGTTTCAGGAGATTAACGTTCCAGAAATTCAGATAC 309
OY 106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTyrLeuGln 125
DB 310 TCTACCAAGCGATCATTTGACGCTAACGCT--AAAGAAACATGCTTCTTCATAGGAA 366
OY 126 IleTyr-----AspGlnGluIleGlyLeuLeuSerGlyLysAla 139
DB 367 AGTTATGTCGACAAATCAAAAGAAACAAATAATTAAGACACTTATGCTGAGACGCT 426
OY 140 GlnLeuAsnGluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMet 159
DB 427 GAGATTAAACACACACAGT-----GTTAATATCTCTCTT 459
OY 160 AspAsnGlyHisPheAlaAsnAspPrometArgTyrAsnGlnGlyTyrProTyrAsn 179
DB 460 GAT-----TCAAAAGCATTCATTAACCAAGGTAACCTTACACACTA 504
OY 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
DB 505 TTGACACCTGTTTATGAAAAGTAACCAACGAGCAACATCTTTAGTCAACATGCA 564
OY 191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTyrProLeu 210
DB 565 GCTACAGAGATGCTGCTACTGCAACTGCTCAATATGAAATATCATTAATACCT-- 621
OY 211 GlnGlyGluGlySerPheAspTyr----- 218
DB 622 ---AACAAAGGGTTGAAAGACTACCTTAGACACTAAGCTCAAAATAACCATATTTCAC 678
OY 219 HisAlaGlySerLeuValGlyAsnTyrSerGlyThrPheGlyGlnMetTyrAspTyrIle 238
DB 679 CATCTTAAGACTG-----TTTGACGCTATCTCTACTGACAAATCAACTGGAAC 729
OY 239 Asn---MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyr 257
DB 730 AACATCTACTACTACTATAGCGGAGAGATCTTAAGCTTCAAAAATG-----GCGATT 783
OY 258 AlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySer 277
DB 784 TCAGAAATGATGCTGATGTTGATTTCTGAGACATGATGATGATGATGATGATGATGATG 843
OY 278 GlyThrTyrSerValTyrValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSer 297
DB 844 TCTGACAGTACGCTCTGCTTCAAAAGACCTTAAAGAAACCTTGGCTTCAACCAATCT 903

```

OY	298	leuglnleuhnlvalarglaaleutyrthrseringlntprhlspawetlleargly	317
	:::	::: :::: ::	
Dd	904	gttccaccmaattmaccsgtagacactttgccaamcaagaattgggaagcacaamtgcacaa	963
OY	318	gluleualaserglyargprovaltytyrallaglaaanaanglnserilleglyltsala	337
	:::	:::	
Dd	964	gaattatcttcamaaccacacagtatactaccmaagtgccggaatgaagtcaggacatccc	1023
OY	338	phevalcyasarglytyralaseraprllyrhpnhsphnastptfgylttprclyal	357
	:::		
Dd	1024	tttgttatcgatggctgcagcagacgttaacctttttcacacatggttaactgtgggtggggctga	1083
OY	358	valserannglyrhetlytleuthrileulseerprothrserleagllylegllyaly	377
	:::	:::	
Dd	1084	gtctctgacggcctcttcgcgtctlgagacacataaaccttcacagctcttggtatcgggtggc	1143
OY	378	glunglylleglyrhnethriletyrglngluilellethrghyglyleglupro	394
	:::	::	
Dd	1144	ggccgcacggcgcttcacacggcttacccaacaaagtcgtgttgagcacaataacct	1194

US-09-221,017B-423
Sequence 423, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PINE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P11182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P11546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 423:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

```

? ANTI-SENSE: UNKNOWN
?
? ORIGINAL SOURCE:
?
? ORGANISM: PORPHYROMONAS GINGIVALIS
?
? FEATURE:
?
? NAME/KEY: misc_feature
? LOCATION: 1...566
?
US-09-221-01/B-423

```

Alignment Scores:

```

rev. no.:      6e-35
Score:         417.00
Percent Similarity: 62.508
Best Local Similarity: 48.308
Query Match:   4.408
DB:            4
Length:        566
Matches:       85
Conservative:  25
Mismatches:    64
Indels:        2
Gaps:          2

```

OS-10-030-330-1 (1-843) X OS-09-221-017B-423 (1-566)

Oy 276 GIGYsergIyThrTYrSerValTYrValValGlyAlaLeuArgAsnAsnPhenArgTYrLys 299
 Db 1 GCGAGTGGCACTTCTCCATATTTGGTAAGCGCCCTCGGGAGACCTTTCATTACACG 60
 Oy 296 ArgSerLeuGlnLeuHisValArgAlaLeuTYrSerGlnGluTYrPHisAspMetIle 315
 Db 61 AAATCGCTTGCTTACATCCATCCGCTCTCTCTCTCCCGGTAAAGAAAGAAAGACATGATT 120
 Oy 316 ArgGlyGluLeuAlaSerGlyArgProValTYrTYrAlaGlyAsnAsnGlnSerIleGly 335
 Db 121 CGTAAAGACTGGCAGAAACACAGACCTGCTATTATATCCGGTGCACAGCATCGATGGCA 160
 Oy 336 HisAlaPheValCysAspGlyTYrAlaSerAspGlyThrPheHisPheAsnTPGlyTyr 355
 Db 181 CATGCTTTGCTTGGCATGGATCGAACCAGACGAGAACCTTCACCTTCAACCTGGGATGG 240
 Oy 356 GlyGlyValSerAsnGlyPheTYrLysLeuThrLeuLeuSerProThrSerLeuGlyIle 375
 Db 241 GGTGGCATANGTACCGGATATTCATTCATTCATCTCAATCAACCGGGTCTCGCTGGCAC 300
 Oy 376 GlyGlyGlnGlyTYrLeuPheThrIleTYrGlnGluIleIleThrGlyIleGluProIle 395
 Db 301 GCGCGAGCGGATGAGGAGCTACTCTACTGACCAAGAGGTTGTATAGAGCATTAAGCCGGCC 360
 Oy 396 LysThrProAlaGluAlaGlyThrAspIleAlaLeuProIleLeuAlaLeuLysAspIleGlu 415
 Db 361 ACCATATGAAGTC---CCGGTATCTGTACCGGATCCGACCATCCGCTTATGATGATGCAA 417
 Oy 416 AlaGluTYrLysSerGlnSerGlyLeuAsnValGlyTYrSerIleTYrAsnThrGlyGlu 435
 Db 418 ---CACCAATATGCTGATGAGAACCTTACCTGACCTGATGTAATAATCAAGACTACTCCACA 474
 Oy 436 GluGlnSerAsnLeuAspLeuGlyTYrTYrAsnLysAsnLysAspGly 451
 Db 475 TATGACAGGGAGTGAATTTGGCTTATGCGCTTGACGCTTCCCAATGGA 572

RESULT 6
US-09-221-017B-129/c
Sequence 129, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows

Db	111	TTCCCGGAGGTGATCGGATNTGCTTTCACAGGGGCACTTGCAATCGGGCCGCTATACGGAC	523
Qy	119	AsnLeuArgMetTrpLeuGlnIleTyrAspGlnGluIle	131
Db	51	AATCTCAGAGGTGGCTCAAGGTTATGACGCTGAATG	13
RESULT 7			
	US-09-221-017B-37		
	Sequence 37, Application US/09221017B		
	Patent No. 6444799		
	GENERAL INFORMATION:		
	APPLICANT: Ross, Bruce C.		
	TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF		
	NUMBER OF SEQUENCES: 1120		
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE: MORRISON & FOERSTER		
	STREET: 755 PAGE MILL ROAD		
	CITY: Palo Alto		
	STATE: CA		
	COUNTRY: USA		
	ZIP: 94304-1018		
	COMPUTER READABLE FORM:		
	MEDIUM TYPE: Diskette		
	OPERATING SYSTEM: IBM Compatible		
	SOFTWARE: FastSeq for Windows Version 2.0b		
	CURRENT APPLICATION DATA:		
	APPLICATION NUMBER: US/09/221,017B		
	FILING DATE: 23-DEC-1998		
	CLASSIFICATION:		
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER: PP1182		
	FILING DATE: 31-DEC-1997		
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER: PP1546		
	FILING DATE: 30-JAN-1998		
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER: PP2911		
	FILING DATE: 09-APR-1998		
	PRIOR APPLICATION DATA:		
	APPLICATION NUMBER: PCT/AU98/01023		
	FILING DATE: 10-DEC-1998		
	ATTORNEY/AGENT INFORMATION:		
	NAME: Monroy, Gladys H		
	REGISTRATION NUMBER: 32,430		
	REFERENCE/DOCKET NUMBER: 27340-20021.00		
	TELECOMMUNICATION INFORMATION:		
	TELEPHONE: 650-813-5600		
	TELEFAX: 650-494-0792		
	TELEX: 706141		
	INFORMATION FOR SEQ ID NO: 37:		
	SEQUENCE CHARACTERISTICS:		
	LENGTH: 656 base pairs		
	TYPE: nucleic acid		
	STRANDEDNESS: double		
	TOPOLOGY: circular		
	MOLECULE TYPE: DNA (genomic)		
	HYPOTHETICAL: NO		
	ANTI-SENSE: UNKNOWN		
	ORIGINAL SOURCE:		
	ORGANISM: PORYPHYROMONAS GINGIVALIS		
	FEATURE:		
	NAME/KEY: misc feature		
	LOCATION: 1...656		
	US-09-221-017B-37		
Alignment Scores:			
	Pred. No.:	1.81e-07	656
	Score:	158.00	53
	Percent Similarity:	41.95%	Conservative: 46
	Best Local Similarity:	22.46%	Mismatches: 101
	Query Match:	3.56%	Indels: 36
	DB:	4	Gaps: 7

Db	1047	GTGTTAGACAAATAT	-----	ACGATTCAGACGGCGGAGAGTGTACCTTTTCACATCCG	1100
Oy	245	AspleuaspbnLeuthInSerGlnValasp	---AlaTyrAlaThrLeuMetcArgasp	263	
Db	1101	ACMAAAGTTAATGCTAATATATAGCAAAATTAAGTAACTGATCTTATGATTAAGTCTCT	1160		
Oy	264	ValSerAlaSerValSerMetSerPheTyrGlnAsnGlySerGlyThrTyrSerValTyr	283		
Db	1161	GCTATTAAGTCTCTGGAGCACTTAATTAATCTCTCTAT	---GATGATGTTTATTCACAAAT	1217	
Oy	284	ValValGlyAlaLeuArg	---AsnAspPheArgTyr	294	
Db	1218	ATTGTACCTGCTTATTCAGATTAAATACAACTGTTACGCTACCTGAAGCGCAAGTTTGG	1277		
Oy	295	---LysArgSerLeuGlnLeu	-----HisValArgAla	304	
Db	1278	GCAACCTATGGCTTCAAACTATTCAGAGCAGAGTACTTCATCTCTCTGCTG	---HisValArgAla	1334	
Oy	305	LeuTyrThrSerGlnGluTyr	-----HisAspMetIleArgGlyGlnLeuAla	320	
Db	1335	GTAATTACTGGGATGATTAATGAGCACTACAACTAAAGCCGTTCAAGTTCATATTC	1394		
Oy	321	SerGlyArgProValTyrTyrAlaGly	-----	329	
Db	1395	AAGGGA	---ACGTAATACCTGCTGCGTTCGTTCAAAAGATACAGTACATATTAAGTT	1451	
Oy	330	-----	AsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyr	343	
Db	1452	ATTGCGTAATTTGGAGAACGACCAAGCA	-----GTTCTTAATTTCTAT	1496	
Oy	344	AlaSerAspGlyThrPheHisPheAsnTyrProGlyTyrGlyVal	---SerAsnGlyPhe	362	
Db	1497	TATTTAAGTCTACCTATTAAGGGTGAAGTAGATTGGAGAGCACTGATATGACGTGGTT	1556		
Oy	363	TyrTyrLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGluGlyIleGlyPhe	382		
Db	1557	ATTAGTTGCTTACAACTTCCCAACCAACCTTAAGTTGCT	1598		
Oy	383	ThrIleTyrGlnGluIleIleThrGly	-----IleGluProAla	395	
Db	1599	ACTATATACGATTCACAAATATTAATTCAAAATTTACAGCTCCATTACTATATGATCCCTAC	1658		
Oy	396	LysThr	-----	397	
Db	1659	AAGATGTATAGTTTCCAAAGAAAGTGACAGACGACAAAGTAGCAAAATATCCGCTC	1718		
Oy	397	-----	-----	397	
Db	1719	ATTGCTCAATGGTCAGAGATGAAACCACTAAAGTATATTCGAAAAATCTATATGCT	1778		
Oy	398	-----	ProIleGluAlaGlyThr	403	
Db	1779	ACTCAGGTTTGGACGACTAAATTTGGAAACAAGAGTGGGATGGTTTGACTATTTCAT	1838		
Oy	403	-----	-----	403	
Db	1839	GACCAAGCTGATAAATTTAATTAACAAGATTTTGGCCGCGAGGTTCAAAAATACA	1898		
Oy	404	-----	AspAlaLeuProIleLeuAlaLeuLys	412	
Db	1899	CTTGAAATGGTACTCCAGCTACAGCTGTAAAGACTACTTATATCTACAAAGAAAGTTC	1958		
Oy	413	-----	AspIleGluAlaGlyTyrTyrLysSerIleSerGlyLeuAsnValGlyTyrSer	429	
Db	1959	AAGATAGTGAATGATCATTTGTGATGATACAGATACCTAGCGAAAAAATTTGTAATTCA	2018		
Oy	430	IleTyrAsnThrGlyGluGlu	---GlnSerAsnLeuAspLeuGlyTyrArg	445	
Db	2019	GTTGTAGATCTCTCAATAGCTCTCTTGCGACAGAGTATATACAGATGTGGACCGTAGA	2078		
Oy	446	-----	LeuAsnLysAlaAspGlyGluVal	453	

Db	2079	CCAGCAGCCTTGTTGGTGGCTGTGATGGGACAGCTACTCTTACAAAGAAAGTTAAAGTCGAT	2138	-----IleGluVal	456
QY	454	-----	---	---	---
Db	2139	TCAGCTAAGACACCGGTACAGTAGTTCACAGTACGACACACTGTTAAGTATGTTTACGAA	2198	---	---
QY	457	LysThrIserSerIleAsnIleSerTrp-----	-----	---	---
Db	2199	AAAGCTGTACGCTTAATGTTAACTCTGTTGACATCAATGGTAAAGTAATCAAAAGCTCTC	2258	---	---
QY	466	-----TyrGlyTyr-----	-----	---	---
Db	2259	GTTTCAGATGAAAAAGATGCGAAACCTGGTTACAAATTGATTAACGACTGGATCAGAAA	2318	---	---
QY	469	-----GlyGluHisPro	472	---	---
Db	2319	TTAGCTTCACACTTTTGAAGGCAAGAAATACAAACTTGTCCCTGCTGGTGAATATCCG	2378	---	---
QY	473	-----GluSer	474	---	---
Db	2379	GTTGGTAAAGTTGGCAGGAAATTACTGATTGAAGTTGCTAATATATCTCCGAAAGCT	2438	---	---
QY	475	PheserIleuAlaProAsnGlnLeuSerGlnGlyIleAsn---ThrIleThrLeuLeuTyr	493	---	---
Db	2439	ATTGACCCAAACACAGCGCAAAATGTAAGCCGGTGTATTACAAAGAAAGTACTAGTCTAT	2498	---	---
QY	494	ArgArThrGlyThrGluGlnTrpGluProValArgHisIalaGlnGlyTyrValAsn	513	---	---
Db	2499	AGACGAGTACAGGCTTCTGTAGTTGTAATTACAAAGATACAGAAAGTTAAAGTATGATTTAA	2558	---	---
QY	514	SerIleLysValAsn-----ThrThrAspPro	522	---	---
Db	2559	GATCCAGAAACGAGTGTCTGATGCACCGGTTGGAGATGCTTATACTACAACTGCACAAAG	2618	---	---
QY	523	Asn---AsnValValValThrValAspAsnAsn-----	532	---	---
Db	2619	AAACCAACGAAATCATCATCAAAAGATGGATCAGCGTATGTTCTTGTCCATCTAAGACA	2678	---	---
QY	533	-----GluGlyLysLeuSerIle-----	538	---	---
Db	2679	GATGGTACGAAAAATGTAAAGTATTCGAAAGAACATCACAGAACTTATGTTATTCACAG	2738	---	---
QY	539	-----ValProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSer	553	---	---
Db	2739	AAAGTTGCMAACTGGATTCACAGAG-----ATTCCAAATGTACACAGAAACAGACCGTCCA	2792	---	---
QY	554	ThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThrProValAlaPhe	573	---	---
Db	2793	AAAGTACTTTCACCATTTGACCCCAACAGACGACGACGACGCCAATTCGATCCA-----	2843	---	---
QY	574	AlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTyrProValMetAlaGlu	593	---	---
Db	2844	-----ACGACACACAGGAACAAATGGCGAGGTCCCAATATTCCTTAC-----	2885	---	---
QY	594	ValProGlyGlySerSerAsnTyrProValValTrpSerLysAspValLeuThrLeuSer	613	---	---
Db	2886	GTTCCAGGA-----TATACACCGGTGATCTCTAAGGATACACGCCCTTGAAA	2933	---	---
QY	614	GluGlyAspTyrThrLeuTrpTyrAlaPheSerIleAsnAsnGlnLysAspGluTrpLys	633	---	---
Db	2934	CCAAATTGAT-----CCAAATGATCCAGGTAAAG	2960	---	---
QY	634	LysIleGlySerVal-----SerValLysThrProThrGluTyr	646	---	---
Db	2961	-----GGTATTATTCACCAACCAACACCAAGAAATTCAGAGTGTGATTACACCAATTCCTTAT	3014	---	---
QY	647	ThrHisProLeuPheGluValGlyHisAsnGlnThrSerThrTyrThrLeuAspMetAla	666	---	---
Db	3015	GTT-----CCAGTTAAAAAAGTCGTAACTAACAC-----GTTGATGAAGAG	3056	---	---
QY	667	HisAsnArgValLeuPro-----AspPheThrLeuLysAsnLeuGlyLeuPro-----	682	---	---
Db	3057	GGTAAACCTTAATGGACCGCAAGAGAGGAACCAAAACCAATCAATGCCAGGTAC	3116	---	---

QY 276 -----glyserglythr----- 279
DB 816 TACATGACGCCCTTACGCAATTAGGTGGCGGTACAGCTTTTCTCTCTGCA 875
QY 280 -----TyrSerVal----- 284
DB 876 AACATGTTGGCTAATACGCTTTCATGATGATGATTTTTCGTATGATACCGCTTA 935
QY 285 ---ValGlyAlaLeuAlaGlyAsnGlyPheArgTyrLysArgSerLeuGlnLeuHisValArg 303
DB 936 GGTATGTTGGCTGCGAATGACTGCGAGACTATTTCAAAGGTAC-----GTTAAC 983
QY 304 AlaLeuTyrThrSerGlnLutPheHisAspMetIleArgGlyGlnLeuAlaSerGlyArg 323
DB 984 GGCATATTCCGCATGAGCGCGCTGCATGATCATCCATACCAAGACTATGATACGCC 1043
QY 324 ProValTyr-----TyrAlaGlyAsnAsnGlnSerIleGly 335
DB 1044 CCACCAATGCGCTTCGATATCCGTTTATGCGCTATACCGCTCATATCCGCGATTAAGCC 1103
QY 336 HisAlaPheValCysAspGlyTyrAlaSerArgTyrPheHisPheAsnTrpLysTrp 355
DB 1104 GCCAAGCTGATATATGACAGCTATATGATGATATGCTTTGTTTAAATCTGAT--- 1160
QY 356 GlyGlyValSerAsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIle 375
DB 1161 -----AAGCTG-----CACTGCATCTCTGTCGCG 1184
QY 376 GlyGlyGlnGlyIleGlyPheThrIleTyrGlnGlnIleIleThrGlyIleGlnProAla 395
DB 1185 GCGACCGTGTGCTTAACCTATACCTCCGCTCTGTCGAGATGGGATC----- 1235
QY 396 LysThrProAlaGlnAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGln 415
DB 1235 ----- 1235
QY 416 AlaGlnTyrLysSerGlnSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyLeu 435
DB 1236 ---GATTAACGCT-----CATGCTACCGGTAAAT 1259
QY 436 GlnGlnSer-----AsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyLeu 452
DB 1260 GAAATGATCTCTTACTACTGACAGTCCGTTATCAGTTGATTAACCAATATTTTGTGAG 1319
QY 453 ValIleGlnValLysThrSerSerIleAsnIleSerTyrGlyTyrGlyGlnHisPro 472
DB 1320 CAATTTGAACACAG-----TATGTTAAACGAGTTA 1349
QY 473 GluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeu 492
DB 1350 AGAACATTATACGACAGCGCTTACAGTCTGGTTCAGCTTAATACCAATATTTTGTGAG 1409
QY 493 TyrArgArg-----ThrGlyThrGlnGln 500
DB 1410 TACAAGAACGAGATATTCTTCTCTGAATATTCGCGATGATATTAAGTACGTAACAC 1469
QY 501 -----Trp 501
DB 1470 AGTACGAGAAAGATTCACTGATGTTAAGCAAAATGAGGTCTGATGATATCTCTGCG 1529
QY 502 ---GluProValArgHisAlaGlnGlyIleTyrValAsnSerIleLysValAsnThrThr 520
DB 1530 GATGATAGTCATTAACGACAGTACAGCGCGCTCAGATTCAGCATAGCGGAACCAAGCGCA 1589
QY 521 AspProAsnAsnValValValThrValAspAsnGlnGlyLysLeuSerIleValPro 540
DB 1590 CAAGACTACAGGCTATTTCTCTGCTATGTCMAAGGTGCGAC----- 1634
QY 541 AsnSerPheValAlaAspLeuAsnSerTyrGlnHisSer----- 553
DB 1635 AATATTATTAAGTACGCGCTCGCGCTATGACCGTAATGCAATAGCTTAACAAATGTA 1694

QY 554 -----ThrIleThrValGlnPheAsnSerAspSerProAspGlnIleArgThrProVal 571
DB 1695 CAGCTTACATTAACCGCTTCTGCGAATGGTCAAGTTGTGACACAGGTGGGTAACGAC 1754
QY 572 AlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTrpValMet 591
DB 1755 TTACAGCGGATTAAGACTTCGGCTAAAGCGGATTAACCGCATACCATTTACTATACCGCG 1814
QY 592 AlaGlnValProGlyGly-----SerSerAsnTyrProValValTrpSerLysAspVal 609
DB 1815 ACGGTGAAAAGAAATGGGTACCTCAGCTAATGTCCTGTTTCATTTAAT----- 1865
QY 610 LeuThrLeuSerGlnGlyAspTyrThrLeuTrpTyrArgPheSerIleAsnAsnGlnLys 629
DB 1866 ---ATTGTTCAAGAACTGCMACTT-----GGGCAATATAGTCCAAA 1907
QY 630 AspGlnTrpLysAlaIleGlySerValSerValLysThrProThrGlnTyrThrHisPro 649
DB 1908 ACGGATGCTAACGTAAGCAACCGTAACGTTGAAGTCAAGTACGCCA----- 1955
QY 650 LeuPheGlnValGlyHisAsnGlnThrSerThrTyrThrLeuAspMetLanHisAsnArg 669
DB 1956 -----GGACAGTCTGCTGTCTGCTAAACCGCGAGATGAGTCA----- 1997
QY 670 ValLeuProAspPheThrLeuLysAsnLeuGlyLeuProPheAsnGlyLeuValVal 689
DB 1998 -----GCACTTAATCCAGTGGCTTAAT 2021
QY 690 ValPheArgGlnThrGlnSerSerGlySerLeuThrAlaAlaGlnLutThrValHis 709
DB 2022 TTTTGTGATCAACCAACGACGATTAAGATGAT---AAGCTGATTAAGCAACTGCA 2078
QY 710 IleLysGlnGlyLeu-----ThrPheValTyrLysProValValGlnGly---Pro 725
DB 2079 GTAGCAATGCTAGAGGATCTTAATTAATCTGTAATGTAATGTAATGTAATGTAATGTAAT 2138
QY 726 IleProAspGlySerTyrArg---AlaThrLeuHisAlaPheValAsnGlyGlnGln 744
DB 2139 GTTAATTAATCAATCCGTTATTCATTCACAAACATTTGGGATGTCAACGGTAAGTCA 2198
QY 745 LeuTyrLeu-----Lys 748
DB 2199 ACGCAAGCAACCAAGGAAATGATGCTGCGACGATTAACATACTTCCAGTTCCGCG 2258
QY 749 GlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaValGlnAlaIleGlnSer 768
DB 2259 GGTAAAGCGAGTGTAGTGCAGACAGTCACTGATGAGGGCTGAGGTTAAACGACGAGTGC 2318
QY 769 Ser-----GlnGlnIleArgVal----- 774
DB 2319 ACTTTTGTGATGAACGTGAATATGACAAACAGTTGATATTTGTTAAATATGTCACA 2378
QY 775 -----PheProAsnProAlaArgAspTyrValGlnIleSerAlaProCys----- 789
DB 2379 GCGAGATGCCATAATATTGCTGCAATATGCTGATTAACGTAACCAAGCGGTGCT 2438
QY 790 -----IleProGlnGlnThrSerIleIleLeuPheAsnLeuSerGly 803
DB 2439 GATGTAATATTCATGATGATTCAGAAATATGAGTCAAGTCACTGTCTCATACATCAGGG 2498
QY 804 LysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeu 823
DB 2499 AAAGTCACTTTAAT-----GTTAAAGGAGAGTGC-----GTAATTAAGGCC 2540
QY 824 ProAsnGly-----AlaTyrIleLeuLysValAspGlyTyrThrThrLys 838
DB 2541 ACATCTGGGATTAACCAACAGTAAAGTAACTATTAAGCAACCGCTGATATGATGAATA 2600
QY 839 IleAsn 840
DB 2601 GTGGAT 2606

RESULT 11

Db 1235 ----- 1235
Qy 416 AlagutyllysSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlu 435
Db 1236 ----GATTACCGT-----CATGGTACGGGTAAT 1239
Qy 436 GluGlnSer-----AsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlu 452
Db 1260 GAAATGATCTCTTACATGCAATGCACTCCGTTATGAGTTGTAATGCTGCTCAG 1319
Qy 453 ValIleGluValLysThrSerSerIleAsnIleSerTyrGlyTyrGlyGlnHisPro 472
Db 1320 CAATATGACCCACAG-----TATGTTAACGAGCTTA 1349
Qy 473 GluSerPheSerLeuAlaProAsnGlnLeuSerGlnLysIleAsnThrIleThrLeu 492
Db 1350 AGACATTTATCAGGACCGCTTACGATCTGTTACGGTAATACAAATATTTATCTGAG 1409
493 TyrArgArg-----ThrGlyThrGluGln 500
1410 TACAGAGACAGGATATCTTCTCTGATATATCCGATGATATTAATGGTACTGAACAC 1469
Qy 501 -----Trp 501
Db 1470 AGTACGAGAGATTCAGTTGATCGTTAAGACAAATACGCTCGATCGTCTGG 1529
Qy 502 ----GluProValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsnThr 520
Db 1530 GATGATATGTCATTCACGACGTACGGGCTGACATTCAGCATAGCGAAGCCAAAGCGCA 1589
Qy 521 AspProAsnAsnValValValThrValAspAsnAsnGlnGlyLysLeuSerIleValPro 540
Db 1590 CAAGACTACACGGCTATTTGCTGCTATGTGCAAGCTGGCAGC----- 1634
Qy 541 AsnSerPheValAlaAspLeuAsnSerTyrGlnHisSer----- 553
Db 1635 AATATTTATTAAGTACGCGCTCGCCTATGACCGTAACTACGCTAACAAATGTA 1694
Qy 554 ----ThrIleThrValGlnPheAsnSerAspSerProAspGlnIleArgThrProVal 571
Db 1695 CACCTTACTATTTACCGTCTGTCGTAAGTGTCAAGTTGTCACACGAGTGGGATACCGAC 1754
Qy 572 AlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTyrValMet 591
Db 1755 TTACGGCGGATTAAGACTTCGGCTAAGCGGATTAAGCCGATACCTACTATTACCGCG 1814
Qy 592 AlagluValProGlyGly-----SerSerAsnTyrProValValTyrSerLysAspVal 609
Db 1815 ACGGTGAAAAAGATGGGTAGCTCAAGCTAATGTCCTGTTCATTTAAT----- 1865
610 LeuThrLeuSerGlnGlyAspTyrThrLeuTyrThrLeuTyrPheSerIleAsnAsnGlnLys 629
Db 1866 ----ATTGTTTCAGGAACTGCACACTCT-----GGCGCAAAATAGTCCCAA 1907
Qy 630 AspGluTyrLysLysIleGlySerValSerValLysThrProThrGluTyrThrHisPro 649
Db 1908 ACGGATGCTAAGCGTAAGCAACCGTAACTGAAGCGATACCGCA----- 1955
Qy 650 LeuPheGluValAlaLysAsnGlnThrSerThrTyrThrLeuAspMetAlaHisAsnArg 669
Db 1956 ----GGACAGGTCTGCTGCTGCTAAACCGCGAGATGAGATGCA----- 1997
Qy 670 ValLeuProAspPheThrLeuLysAsnLeuGlyLeuProPheAsnGlnGlyLeuValVal 689
Db 1998 ----GCACCTAATGCGCACTGGCGTATA 2021
Qy 690 ValPheArgGlnThrGlnSerSerSerGlySerLeuThrAlaGlnGlnThrValHis 709
Db 2022 TTTTGTATCAAAACCAAGCCACATTTACTGAGATT---AAGCTGATTAAGACAACTGCA 2078
Qy 710 IleLysGlnGlyGlu-----ThrPheValTyrLysProValValGlnGly---Pro 725
Db 2079 GTAGCAAAATGCTAAGTACTATTAATATACGTAAAGTATGAAAAACGGTACACCA 2138

Qy 726 IleProAspGlySerTyrArg---AlaThrLeuHisAlaPheValAsnGlnGln 744
Db 2139 GTTATATATCAATCCGTTACATCTCAACAACTTTGGATGTTCAACGGTAAGTCTCA 2198
Qy 745 LeuTyrLeu-----Lys 748
Db 2199 ACGCAACCAACCAAGGGAATGATGCTGTCGACGATTAACACTACTTCACAGTCCGCC 2258
Qy 749 GlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaValGluAlaIleGluSer 768
Db 2259 GGTAAACGCACTGTATGCGACAGTCAGTATGCGGCTGAGTTAAAGCGACTGAGTCC 2318
Qy 769 Ser-----GluGluIleArgVal----- 774
Db 2319 ACTTTTGTATGACGAAATTAAGACAAAGGTTGATATTTATGTTACAAATGTCAGA 2378
Qy 775 ----PheProAsnProAlaArgAspTyrValGluIleSerAlaProCys----- 789
Db 2379 GCGGATGGCTATATTTGGCTGCATATGTCAGTTTAACTGAAGCAAGCGGTGT 2438
Qy 790 -----IleProGlnGluThrSerIleIleLeuPheAspLeuSerGly 803
Db 2439 GATGTACATATTCATGATGATTCAGAAATACAGTATCGGACTGTCATGATCAGG 2498
Qy 804 LysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeu 823
Db 2499 AAGTCACTTGAAT-----GTTAAAGCAAGTGTCTC-----GTTAATTAAGCC 2540
Qy 824 ProAsnGly-----AlaTyrIleLeuLysValAspGlyTyrThrThrLys 838
Db 2541 ACATCTGCTGATTAAGCAAAAGTAACTTACACTTAATAAGCAGCGTGTATGATATAA 2600
Qy 839 IleAsn 840
Db 2601 GTGGAT 2606

RESULT 12
US-08-277-231A-1
Sequence 1, Application US/08277231A
Patent No. 5643725
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP p111n
Patent No. 5643725
TITLE OF INVENTION: Structural Genes and The LKP p111 Operon of No. 5643725tpa
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277, 231A
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 32,542
REFERENCE/DOCKET NUMBER: ACC94-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 1:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 9432 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1882..2532)
NAME/KEY: CDS
LOCATION: 2854..3630
FEATURE:
NAME/KEY: CDS
LOCATION: 4016..6238
FEATURE:
NAME/KEY: CDS
LOCATION: 6259..6873
FEATURE:
NAME/KEY: CDS
LOCATION: 6955..8265
NAME/KEY: CDS
LOCATION: 8395..9340
US-08-277-231A-1

Alignment Scores:
Pred. No.: 0.00115
Score: 140.00
Percent Similarity: 33.15%
Best Local Similarity: 21.24%
Query Match: 3.15%
Matches: 196
Conservative: 110
MisMatches: 326
Indels: 291
Gaps: 46
US-10-030-330-1 (1-843) x US-08-277-231A-1 (1-9432)

OY 22 AlaProValThrLysGluArgLeuSerLeuAlaArgLeuAlaLeuArg----- 38
DB 4571 GCCCAGTGTGGTGGCGTGGCAACACAAACGCCAAAGTCAGCATCAACAAATAGGC 4630
OY 39 -----GlnValSerLeuArgMetGlyGlnThrAlaValSerAspLysIleSer 54
DB 4631 TATACGATTTATCAACACCGTCCCGCAGGCGCTTTCGTGATTAACGATTTGTATGCC 4690
OY 55 IleAspTyrValTyrArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGluGly 74
DB 4691 ACCGGTTAT-----AGCGCGCATTTAACGGTGAATC---CAAGAAAGTGATGCT 4738
OY 75 SerProAlaTyrPheTyrValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuValAla 94
DB 4739 AAAGTGGCGTCATTTATTGTGCG-----TTTCTCAATCTTGCC 4777
OY 95 AlaAspAspArgIleProThrIleLeuAlaTyrSerProIleGlyArgPheAspMetasp 114
DB 4778 CGGTAAATGCCGTGGGCGCATTTGGCTTATCAATTAAGCTGGCAGCATTAATGCAATTGAC 4837
OY 115 SerMetProAspAsnLeuArgMetTyrPheGlnIleTyrAspGlnGluIle-----Gly 132
DB 4838 AGC-----CGCACCTTTGATGAACGTGCTGTTTCAAGGC 4870
OY 133 LeuIle-----LeuSerGlyLysAlaGlnLeuAsnGluGluIleLeuArgThrGlu 149
DB 4871 GTGTGGCAATGTGTTAACTAATCAATCACTACGCTGAATCAAGCCTGTTTATACAGCT 4930
OY 150 GlyAlaProAlaGluValAlaIleAlaLeuMetAspAsnGlyHisPheAlaAsnAppromet 169
DB 4931 CATATAGTGTGAGGCGCTTT-----GCTTTTGCTTTAAATACGCCGAT 4975
OY 170 ---ArgTyrAsnGlnGlyTyrProTyrAsnAsnLysGlu---ProLeuLeuProAsnGly 187
DB 4976 GGGCGGCTTCTGTGATGCACCTGTGTCGACGCGAATTTCCGCTAAACATGTGACG 5035
OY 188 AsnHisAlaTyr-----ThrGlyCysVal 195

DB 5036 AAAAAGGCTACAGCTTCGACGCGCACTTATAGTATTACTTAATGAAGTGGCACCAAT 5095
OY 196 AlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTyrProLeu----- 210
DB 5096 ATCAGCTGGCAGCC-----TATCGCTATCTTCACGCGGATTTTACACCTTAAGCAGC 5149
OY 210 ----- 210
DB 5150 ACCATGTGCTTACCGCAGCTTCACAGACATTTAGCGGTATTTGCTGAATTTAC 5209
OY 211 GlnGlyGluGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGlyThr 230
DB 5210 CGCCCAAAAATCACTTCAAGTGAAGTAAAGCCAAATCTGGGAAATGGGAAATCTC 5269
OY 231 Phe-----GlyGluMetCysTyrAspTyrPheAsnMetProGlyAsnProAspLeuAspAsn 248
DB 5270 TATCTTTACAGGACCAACCTATATATTTATTTGGGAAAAACGTGCGCAGCAAT----- 5317
OY 249 LeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerVal 268
DB 5318 ---ACGCAATATCAAGTT---GCCATTCAAACAGCTTCACATTTTAATTAAGTCTGTA 5371
OY 269 SerMetSer-----PheTyrGluAsnGlySerGlyThrTyrSerValTyrValVal 285
DB 5372 AACCTCTACAGAGATTTATGATTAAGAAAACGGGCAACGTGACACACACATTTATTAAGT 5431
OY 286 GlyAlaLeu-----ArgAsnAsnPheArgTyrLysArgSer----- 297
DB 5432 CTCAGCGCTGCCATTAAGGCGAATACCATTTCTGCAGATGATGTTATCTCCAGGGGTAC 5491
OY 298 -----LeuGlnLeuHisValArgAlaLeuTyrThrSerGlnGluTyrPheAspMet 314
DB 5492 GATTTTACCAACAGCAGTGGGTAAAGGCTCTTT-----GGTGAACGTATCAATG 5545
OY 315 IleArgTyrGluLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnIle 334
DB 5546 AGTATGTGATTAACGCTTCA-----CGCAATATCAAA----- 5578
OY 335 GlyHisAlaPheValCysAspGlyTyrAlaSer---AspGlyThrPheHisPheAsnTyr 353
DB 5579 -----GGCTATCGCAGTTATGACGGTAACTTTCGATTAACAT 5617
OY 354 GlyTyrPheGlyValSerAsnGlyPheTyrGlyLeuThrLeuLeuSerProThrSerLeu 373
DB 5618 AGCATGTGATGATAC---CGTCTCTTATTCACGTATAGCCTCAAAAATCGCTCCATC 5674
OY 374 GlyIleGlyGlyGluGlyIle-----GlyPheThrIleTyrGlnGlu 387
DB 5675 TCACTGGGCGCAAGCGGTGCTGTCGTGGCGCACAAACAGCGATTAACCTTAAGCCACCT 5734
OY 388 IleIleThrGlyIleGluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuPro 407
DB 5735 GTT-----GGCAAAAGTTTCC 5752
OY 408 IleLeuAlaLeuLysAspIleGluAlaGluTyrLysSerGluSerGlyLeuAsnValGly 427
DB 5753 ATTATTCACGCCAAAGAT---GCCGAGAGAGAAAAGTGATTCAGGTGCCAATGTGACG 5809
OY 428 TyrSerIleTyrAsnThrGlyGluGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsn 447
DB 5810 CTTGATTAATTC----- 5821
OY 448 LysAlaAspGlyGluValIleGluValLysThrSerSerIleAsnIleSerTyrTyrGly 467
DB 5822 -----GGCAATGCGGTATGCTTACACACGCCGATGTGAATTAATTAATGCGT 5872
OY 468 Tyr-----GlyGluHisProGluSerPheSerLeuAlaProAsnGlnLeuSer 483
DB 5873 ATCAATTCATCTGATGCGAGGCGAATGTGAATTTGAAGCACATGAACGCCAAATCAT 5932
OY 484 GlnGlyIleAsnThrIleThrLeuLeuTyrArgTyrGlnGlyThrGlnGluGlnTyrPro 503

```

```

Db 5933 CCTCGTCAATTCATAGTATAGTATTCGCCGACGGCAAAAT----- 5980
Oy 504 VALATGHSALAGLNGLYGTYTVALAsnSerIleLysValAsnThrAspProan 523
Db 5981 -----ACAAATGGTGTATTAACTCCTCACTTGGCCAAAT 6013
Oy 524 AsnValVal-----ThrValAsnAsnGluGlyLysLeuSerIleVal 539
Db 6014 GCGGAGCCAGTCCCAATGGCAATCCGCCAAGAGG----- 6058
Oy 540 ProAsnSerPheValAlaAspLeuAsnSerTYrGluHisSerThrIleThrValGlnPhe 559
Db 6059 -----GCATTGTGGCGCATGT----- 6076
Oy 560 AsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGlyAla 579
Db 6077 -----GTCAAGGTGGTGTCTTGTGCTTAATAACTTACCCAG 6115
Oy 580 ThrAlaAspAspValIleSerLeuGlyTrp-ValMetAlaGluValProGlyGlySerse 599
Db 6116 CCAAAAGCGAGTTAATC---GTCAATGGGTGGAGGAGAAAGCAACAAATGCCGTTTC 6172
Oy 599 TAsnTYrProValValTrpSerLysAspValLeuThrLeuSerGluGlyAspTYrThrle 619
Db 6173 CAATATCAAGTGTGATTGG-----ATACGCA 6199
Oy 619 UTPrTYrArgPheSerIleAsnAsnGlnLysAspGluTrpLysIleGlySerValse 639
Db 6200 CAAATACAAAGTCAGATATTCATACCAACCGCAAAATTAATATATTAAGAGCATTTA 6259
Oy 639 rVal-LysThrProThrGluTYrThrHisProLeuPheGluValGlyHisAsnGlnHis 659
Db 6260 TGCAAAAACACCCAAAAATTAACC---GCCGTTTTC-----CATCAAAAT 6304
Oy 659 eTrhTYrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnL 679
Db 6305 CCATGCTGCTTGTAGT----- 6321
Oy 679 eugLYLeuProPheAsnGly-GluLeuValValPheArgGlnThr-GlnSerSerse 698
Db 6322 --GGAGCAAAATTAAGTGAGCAAAATTAAGTGCGTCAAAATGCTTACGTTTCAACGTC 6379
Oy 698 rGlySerLeuTrpAlaAlaGlnGluThrValHisIleLysGlnGlyGluThrPheValY 718
Db 6380 TGGCTCTGCTCTGCTGCGT----- 6398
Oy 718 rLysProValValGluGlyProIleProAspGlySerTYrArgAlaThrLeuHisAlaPh 738
Db 6399 -----GGCTCT-----GCTTGAATGCTTGT-- 6419
Oy 738 eValAsnGlyGlnGlnLeuTYrLeuLys-GlyLysArgAsnTYrThrValLysIleV 758
Db 6420 -----GGCACTGCTGCTTATGCTTACGATGGAGTGAACCTTCAAGGGGAGATTT 6472
Oy 758 al---AsnGlyThrAlaValGluAlaIleGluSerSerGluGluLeuValPheProa 777
Db 6473 TAGATGATGGCACTGT-----AAATGAAACACACGCCAAATTCGACGGTTACCC 6526
Oy 777 snProAlaArgAspTYrValGluIleSer-----AlaProCysIleProGlnGluT 794
Db 6527 TCCCAACAGTGGAAAGCTAATTAAAGCAGCAGGAAAGCCGCCGCTGTCCTT 6586
Oy 794 hSerIleIleLeuPheAspLeuSerGlyLysIleValMetLysAsnSerLeu-----S 812
Db 6587 TTTTCATCAGCTTAAGAAATGCAATGCAGATGATGTAAGAAAGCTAATCTGCTATTTA 6646
Oy 812 eAlAGlyHisGlyArgMetAspValSerArgLeuProsn----- 825
Db 6647 AAGGGGAGACACACACAGGCAATCTTATCTTCCAAATTAAGCAGCAACGGCAAG 6706
Oy 826 -----GlyAlaTYrIleLeuLysValAspGlyTYrThrThrLysIleAsnIle 841
Db 6707 CCACCAACGTGGGCAATTCAAATTTGTCAAAGCCGATGGCATAGCAGCGCTATCAAGGTG 6765

```

```

RESULT 13
US-08-473-750-4
; Sequence 4, Application US/08473750
; Patent No. 5834187
; Patent No. 5834187
; Patent No. 5834187 5786143
; GENERAL INFORMATION:
; APPLICANT: Green, Bruce A.
; APPLICANT: Brinton, Jr., Charles C.
; TITLE OF INVENTION: Sequence and Analysis of LKP P11n
; Patent No. 5834187
; Patent No. 5834187 5786143
; TITLE OF INVENTION: Structural Gene and the LKP P11 Operon of No. 5834187 5786
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millita Drive
; City: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,750
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,321
; FILING DATE: 19-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Alice O.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: ACC94-02B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (1882..2532)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2854..3630
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4016..6238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6259..6873
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6955..8265
; NAME/KEY: CDS
; LOCATION: 8395..9342
; US-08-473-750-4
;
; Alignment Scores:
; Pred. No.: 0.00115
; Score: 140.00
; Percent Similarity: 33.15%
; Best Local Similarity: 21.24%
; Length: 9432
; Matches: 196
; Conservative: 110
; Mismatches: 326

```


Db 6260 TGCACAAATACCCCAAAATTAACCC---GCCCTTTTC-----CATCAAAAT 6304
 Oy 659 eRTThTThLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnL 679
 Db 6305 CCACGCTACTGTTGACT----- 6321
 Oy 679 euGLLeuProPheAsnGly-GluLeuValValPheArgGlnThr-GlnSerSerSe 698
 Db 6322 --GGAGCAAAATATATAGTGAGCAAAATATATAGGGGCTCAAAATGCTTAGCTTCAATGTC 6379
 Oy 698 rGLYSerLeuThrPalaIaGlnGluThrValHisIleLysGlnGlyGluThrPheValTy 718
 Db 6380 TGGCTCTGCTGCTTGCCT----- 6398
 Oy 718 rLysProValValGluGlyProIleProAspLysSerTyraGalaThrLeuHisAlaPh 738
 Db 6399 -----GGCTCT-----GCTTAGTTGCAATGCT-- 6419
 Oy 738 eValAsnGlyGlnGlnGlnLeuTyrrLeuLys-GlyLysArgAsnTyrrThrValLysIleV 758
 Db 6420 -----GGCACTGCCCTGCTTATGCTTACGATGCGAGATGCACTTTCAAGGAGATTT 6472
 Oy 758 aL---AsnGlyThrAlaValGluAlaIleGlnSerSerGlnGluLysLeuArgValPhePro 777
 Db 6473 TAAGTAGGAGCACTGCT-----AAATTTGAACACAGACACCCAAATGCTACCGGTTAACCC 6526
 Oy 777 snProAlaArgAspTyrrValGluIleSer-----AlaProCysIleProGlnGluT 794
 Db 6527 TGCCACACAGTGGAAACCTATATTAAAGCAGCGAGGCAACGCCGCCCTGTGGCT 6586
 Oy 794 hrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLysAsnSerLeu-----S 812
 Db 6587 TTTTCATCTACGCTTAAAGAAATGCAATGCAATGATGCTATGTAAAGCTATCTGTATTTA 6646
 Oy 812 eTaIaGlyHisGlyArgMetAspValSerArgLeuProAsn----- 825
 Db 6647 AAGGGGAGACACACACACACAGGCAATCTTATCTTCCAAATTAAGCAGGCAACGCCAAG 6706
 Oy 826 -----GlyValATyrrIleLeuLysValAspGlyTyrrThrThrLysIleAsnIle 841
 Db 6707 CCACCAACGCTGGGCAATTCAAATTCATCAAAAGCCGATGCGATAGGCAAGCCTATCAAGTG 6765

RESULT 14
 US-08-477-326-4
 : Sequence 4, Application US/08477326
 : Patent No. 5968769
 GENERAL INFORMATION:
 APPLICANT: Green, Bruce A.
 APPLICANT: Brinton, Jr., Charles C.
 TITLE OF INVENTION: Sequence and Analysis of LKP P11n
 Patent No. 5968769
 TITLE OF INVENTION: Structural Gene and the LKP P11 Operon of No. 5968769lypable
 NUMBER OF INVENTION: Haemophilus Influenzae
 CORRESPONDENCE ADDRESSES: 21
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 STREET: Two Millitia Drive
 CITY: Lexington
 STATE: Massachusetts
 COUNTRY: US
 ZIP: 02173
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/477,326
 FILING DATE: 07-JUN-1995
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/277,231

US-10-030-330-1 (1-843)	US-08-477-326-4 (1-9432)	US-08-477-326-4
0y	22	22
1y	4571	4571
2y	39	39
3y	4631	4631
4y	55	55
5y	4691	4691
6y	75	75
7y	4739	4739
8y	95	95
9y	4778	4778
10y	115	115
11y	4838	4838
12y	133	133
13y	4871	4871

```

OY 150 G1yvalProAlaGluValAlaMetAspAsnGlnHisPheAlaAsnAspProMet 169
DB 4931 CATATGCGTGCAGGCGCTTT-----GCTTTGGTTAAATACGGCAT 4975
OY 170 ---ArgTPAsnGlnGlyTyrProTPAsnAsnLysGlu---ProLeuPheProAsnGly 187
DB 4976 GGGCGCTTTCTGCTGATGCGACATGTCGACAGCTGAATTCGGCTAAACATGTGAC 5035
OY 188 AsnHisAlaLysTyr-----ThGlyCysVal 195
DB 5036 AAAAAGCGGTACAGCTGCAGCGGAGATATGATTAATCAATGAAGAAGTGGACCAAT 5095
OY 196 AlaTPHisAlaAlaGlnLeuMetArgTyrHisSerTyrProLeu----- 210
DB 5096 ATCAGCTTGGCAGCC-----TATGCTATTCCTTCAGCGGATTTTACACTTAAGCAG 5149
OY 210 ----- 210
DB 5150 ACCATTGGCTTAAACCGCACTTTCAGACAATTAGCGGTGCTAATTTGGCTGAATTTAC 5209
OY 211 GlnGlyGlnGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGlyThr 230
DB 5210 CGCCCAAAAATCAAGTTTCAAGTAGTTTAAGCCAAAGCTGGGGAATTTGGGAATCTC 5269
OY 231 Phe-----GlyGluMetTyrAspTyrPheAsnMetProGlyAsnProAspLeuAspAsn 248
DB 5270 TATCTTTCAGCAACAACCTTAATTAATTTGGAAAAACGTGGCAAGT----- 5317
OY 249 LeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerVal 268
DB 5318 ---ACGCAATATCAGAGTT---GCCTATTCAACAGCTCCACATTTTAACTCTGTGA 5371
OY 269 SerMetSer-----PheTyrGluAsnGlySerGlyThrTyrSerValTyrValVal 285
DB 5372 AACCTCACAAGAGTATGATTAAGAAACGGGCAAAAGTACAAACGCAATTAATTAAGT 5431
OY 286 G1yAlaLeu-----ArgAsnAsnPheArgTyrGlyAspSer----- 297
DB 5432 CTCAGCTGCGCATTAAGCGCATTAACCATCTGCGAGATAGATATCTTCCAGTGGTAC 5491
OY 298 ---LeuGlnLeuHisValArgAlaLeuTyrThrSerGlnGluTyrHisAspMet 314
DB 5492 GATATTACCAACAGACTTGGCGTAAATGGCTCTTTT---GCTGAACGCTCATCAATGG 5545
OY 315 IleArgGlyGluLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIle 334
DB 5546 AGTTATGATTAACGCTTCA-----CGCAATTAATCAA----- 5578
OY 335 GlnHisAlaPheValCysAspGlyTyrAlaSer---AspGlyThrPheHisPheAsnTyr 353
DB 5579 -----GGCTATCGCAATGATTAAGCGTAATCTTCCCATTAACAAAT 5617
OY 354 G1yTyrGlyGlyValSerAsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeu 373
DB 5618 ACCATTGGTAGTTAC---CGTGCCTCTTATTCACGTGATACCTCAAAAATCGCTCATC 5674
OY 374 G1yIleGlyGlyGlyIle-----GlyPheThrIleTyrGlnGlu 387
DB 5675 TCACGGGCGCAAGGGGTGCTGTGCGGACAAACAGTATTAACCTTAAGCAACCT 5734
OY 388 IleIleThrGlyIleGluProAlaLeuThrProAlaGlnAlaGlyThrAspAlaLeuPro 407
DB 5735 GTT-----GGCAAAAGTTTTGGC 5752
OY 408 IleLeuAlaLeuAspIleGlnAlaGlyTyrLysSerGlnSerGlyLeuAsnValGly 427
DB 5753 ATTATTACGCCAAAGAT---GCCGCGAGACAAAGATGGAATCAGGTGCACATGTGAC 5809
OY 428 TyrSerIleTyrAsnThrGlyGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsn 447
DB 5810 CTTCATATATTTC----- 5821
OY 448 LysAlaAspGlyValIleGlyValLysThrSerSerIleAsnIleSerTyrGly 467

```

```

DB 5822 -----GCAATGGGTTATGCTTACACCGACCGGCTAATGAATCAATTATATCGT 5872
OY 468 Tyr-----GlyGlnHisProGluSerPheSerLeuAlaProAsnGlnLeuSer 483
DB 5873 ATCAATTCATCTGATGGGAGGCGGATGTGGAAATTTGAAGCCATGAACGCCAAATCAT 5932
OY 484 GlnGlyIleAsnThrIleThrLeuLeuTyrArgArgTyrGlyThrGlnGlnTyrPro 503
DB 5933 CCTCGGCAAAATCAATTAATGCTTAGTAGATTTCCGACGGGCAAAAT----- 5980
OY 504 ValArgHisAlaGlnGlyGlyTyrValAsnSerIleLysValAspThrThrAspProAsn 523
DB 5981 -----ACATGGGTATTTAATCACTTCATTTGCCAAT 6013
OY 524 AsnValValVal-----ThrValAspAsnGlnGlyLysLeuSerIleVal 539
DB 6014 GCGGAGCCAGTCCATGCGATCCACCGCAACAGTACGAGG----- 6058
OY 540 ProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPhe 559
DB 6059 -----GCATTTGTGGCGATGTG----- 6076
OY 560 AsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGlyAla 579
DB 6077 -----GTGCAAGGTGTGTGCTTTTCCGTAATAACTTAACCTTACCAG 6115
OY 580 ThrAlaAspAspValIleSerLeuGlyTyrValMetAlaGlyValProGlyLysSer 599
DB 6116 CCAAAAGCGCAATTAATC---GTCAAAAGGGGTAGGAGAAAGCGCAACCAATGCCCTTC 6172
OY 599 AsnTyrProValValTyrSerTyrAspValLeuThrLeuSerGlyLysArgTyrThr 619
DB 6173 CAATATCAAGTGTATTTGG-----ATAACGCA 6199
OY 619 TyrTyrArgPheSerIleAsnAsnGlnLysAspGluTyrGlyLysIleLysSerVal 639
DB 620 CAATATCAAGTCAATGATTCATGCAAAAGCGCAAAATTAATTAATTAAGAGATTTA 6259
OY 639 rVal-LysThrProThrGlyTyrThrHisPheLeuPheGluValGlnHisAsnGlnThr 659
DB 6260 TGCACAAAACCAACCAAAATTAATC---CGCGTTTTC-----CATCAAAAT 6304
OY 659 eThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuAsnL 679
DB 6305 CCACTGCTACTTGTACT----- 6321
OY 679 euuLysLeuProPheAsnGly-GluLeuValValPheArgGlnThr-GlnSerSer 698
DB 6322 --GGACCAATATATAGTGAGCAAAATTAATGAGTGCACAAATGCTTACGTTTCATGCTC 6379
OY 698 rGlySerLeuThrAlaAlaGlnGlnThrValHisIleLysGlnGlyLysThrPheValTyr 718
DB 6380 TGGCTGTGCTGCTTGGCT----- 6398
OY 718 rLysProValValGlnGlyProIleProAspGlySerTyrArgAlaThrLeuHisAlaPhe 738
DB 6399 -----GGCTCT-----GCTGATGTCATTTG-- 6419
OY 738 eValAsnGlyGlnGlnGlnLeuTyrLysLeu-GlyLysAspAsnTyrThrValLysIleVal 758
DB 6420 ---GGACCTGCTCTTATGCTTAAGTGCAGACAGTGCACCTTCAAGGAGGATTT 6472
OY 758 al---AsnGlyThrAlaValGlnAlaIleGlnSerSerGlnGlnIleArgValPhePro 777
DB 6473 TAAAGTAGGACACTTGT-----AAATTAACACAGACCAAAATGCGAGGTTAACCC 6526
OY 777 snProAlaArgAspTyrValGlnLysSer-----AlaProCysIleProGlnGlnTyr 794
DB 6527 TGCACACAGTGGGAAAGCTAATTAATTAAGCCACGAGGCAAAACCGCGCCCTGTGCTT 6586
OY 794 hrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLysAsnSerLeu-----S 812

```


OY 271 SerPheThyGluAsnGlySerGlyThrTySerValIleValAlaLeuArgAsn 290
 : : : : :
 Db 851 AATTGTTGGGGTGGGGTAGT---TACAACGCTAAATCTCAGGCATA----- 898
 OY 291 AsnPheArgTyLysArgSerLeuGln----- 299
 Db 899 -----AGAACTTATCCGGTGATATCATACGAGATGAGTGGGTCC 943
 OY 300 -----LeuHisValArgAlaLeuTyThrSerGlnIleTrp----- 311
 Db 944 TTGATTTATGCTTACGCTAAACGATATATAT---TCATGCTCAATGGGGTCCCGCTGAT 1000
 OY 312 -----HisAspMetIleArgGlyGluLeuLeuAlaSerGlyArg 323
 Db 1001 GACGGAACATTTACACAGCCCTAGTACCTGGTAAAGAGCTTTAGTAAAGGT--- 1057
 OY 324 ProValTyThrValAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValAspGlyTy 343
 Db 1058 -----GTTACTAGGAGAGAGATCCCAAGAGCGATTTACGTT-----TTT 1099
 OY 344 AlaSer-----AspGlyThrPheHisPheAsnTrpGlyTyThrGlyValSerAsnGly 361
 Db 1100 GCCAGTGAATAGCTGGAGACTGCTGCTGATATTCGAATTACGACGCTATACTAATCC 1159
 OY 362 PheTyLysLeuThr-----LeuLeuSerProHisSerLeu 373
 Db 1160 ATATATTTCTATTTACTTTGGGCTATTTGATCACAAGATCTTACATCTCTTTATTCGAA 1219
 OY 374 GlyIleGly-----GlyGluGlyIleGlyPheThrIleTyGln 386
 Db 1220 GATTGTTCCCGCTCATGCGACAGTACAGTATTTCTTCAAGTTCAAGGATATATTTCTG 1279
 OY 387 GluIleIleThrGlyLe-----GluProAlaIleThrProAla 399
 Db 1280 AGTGAATATCAACGCGAGATGACATAATACCAACGCGTCCGCTGCTCATTA 1339
 OY 400 GluAlaGlyThr-----AspAlaLeuProIleLeuAlaLeuLysAspIle 414
 Db 1340 GGTGCGCGTGTACACTTTGTTACTAGAACCCCAACCACTTCTGGAGACGCTA 1399
 OY 415 GluAlaGlyTyLysSerGlySerGlyLeuAsnValGlyTySerIleTyArgIleGly 434
 Db 1400 CAGTAT-----TTATCATCTTGTCTGCGGTA 1426
 OY 435 GluGluGlnSerAsnLeuAspLeuGlyTyArgLeuAsnLysAlaIle 454
 Db 1427 GGGTAGAAGAAAGACGCTGACGAGATTTGGAGA---GATAGCGCCCTGGGAGAAATAC 1483
 OY 455 GluValIleTyThrSerSerIleAsnIleSerTrpTyGlyTyGlyGlnHisProGlySer 474
 Db 1484 TGTCTATCGC-----TATGCGTTGGTAAATCCATGCCCAT 1519
 OY 475 PheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleLeuLeuTyArg 494
 Db 1520 AAGTTAATTAATGAATGCCAAGACCTGGAGAAATGTTAACGCAACCACTGGTTTACCTG 1579
 OY 495 ArgThrGlyThrGluGlnIleTrpGluProValArgHisAlaGlnGlyTyArgValAsnSer 514
 Db 1580 CCACACA-----TTGATGATTTTCCCAAG 1600
 OY 515 IleTyValAsnThrThrAspProAsnAsnValIleValIleThrValAspAsnGlnGly 534
 Db 1601 TCCACAACTCCACGAGAGACATTAAGATCCGTTATTAACCTA-----TCAGAAAA 1654
 OY 535 LysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyGlnHisSerThr 554
 Db 1655 ACTCTTCAA-----GATGCTTAACCTCAAGATTTGAGCAGCTGACAG 1696
 OY 555 IleThrValGlnPheAsnSerAspSerProAspGlnIleArgThrProValAlaPheAla 574
 Db 1697 GTTACTGTA-----GATATTTGATACGAATTTAGG----- 1726
 OY 575 LeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTyThrValMetAlaGluVal 594

Db 1727 -----GGAAGTACAGCTGCGATTTAATATCA----- 1753
 OY 595 ProGlyGly-----SerSerAsnTyThrProValIleTrpSerLysAspValLeuThrLeuSer 613
 Db 1754 CCAGCGGGGATTAATTTCAAACTTGGCGTGTGAAGCAAGAGATGTT-----TCATCA 1807
 OY 614 GluGlyAspTyThrIleuTrpTyArgPheSerIle-----AsnAsnGlnLys 629
 Db 1808 GAGGGA-----TTCAAGACCTGAGACATTCATGCTGTAGACATTTGGGTGAGAACGGGCTA 1864
 OY 630 AspGluTrpLysLysIleGlySerValSerValIleTyThrProThrGlyTyThrHis 648
 Db 1865 GGTGATTGAAA-----ATCAAGTTTAAGCA---ACAGAAAAATGGACACAGG 1909
 OY 649 -----ProLeuPhe-----GluValGlnHisAsnGlnThr 658
 Db 1910 ATTAGCTTCCACAGTTGAGAGCTGAGAGCTTTGGGGAATTCATTTGATTCATCAAAACA 1969
 OY 659 SerThrTyThrIleuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsn 678
 Db 1970 GAACCTTGTGCTTTGGAAACGATTAAGAGAGAGGTGAAACACCTGCTACGAAGATACC 2029
 OY 679 LeuGlyLeuProPheAsnGlyGluLeuValValIlePheArgGlnThrGlnSerSer 698
 Db 2030 GTATTCACATATTTCTGCGACTTCACTTATTTCCATGACGCGCTACTTCTACATCTTCT 2089
 OY 699 GlySerLeuThrAlaAlaGlnGlnThrValHisIleLysGlnGlyGlnThrPheValTy 718
 Db 2090 ATCTCAAT-----GGTGTGAAGACGTGGCGGCTATCCCAAGACATCTGGAGATACC 2143
 OY 719 LysProValValGluLeuGlyProIleProAsp-----GlySerTyArgAlaThrLeuHis 736
 Db 2144 GATCGTGAATTCGATTCCAAAACACTCTTAAAACTTCTCTCTAGCAAGCAATGCAT 2203
 OY 737 AlaPheValAsn-----GlyGlnGlnGlnLeu-----Ty 746
 Db 2204 TATTTTAAACATATTTTGTATGATGGCGCCACATTTTGGTGTATTTACTTCAATGTTT 2263
 OY 747 LeuLysGlyLysArgAsnTyThrValLysIleValAsnGlyThrAlaValAlaIle 766
 Db 2264 ATCAATTCACAGGAGAGG-----ATCAGAGGTCAAGAGCGGAACGAT 2308
 OY 767 Glu-----SerSerGluGlnIleArgValAlaPheProAsnProAlaArg 780
 Db 2309 GAATTCGATATTCATGATACAGACTGTGACTGACGATTTCTTGGACATTTGGAATCC 2368
 OY 781 AspTyValGluIleSerAlaProCysIleProGlnGlnThrSerIleIleLeuPheAsp 800
 Db 2369 GGAATTCAGG-----CCGGAAGAGTTGAGACTTCGATTTTGAT 2410
 OY 801 LeuSerGlyLysIleValMetLysAsnSerLeuSerAlaGlyIleGly 816
 Db 2411 TTCTCCGATGAGAACCATCTTGCAAGTTGTCTTCAACGAAACGCT 2458
 RESULT 17
 US-09-453-702B-18
 ; Sequence 18, Application US/09453702B
 ; Patent No. 6365723
 ; GENERAL INFORMATION:
 ; APPLICANT: Blattner, Frederick R.
 ; Burland, Nicole T.
 ; Perna, Nicole T.
 ; Plunkett, Guy
 ; Welch, Rod
 ; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
 ; NUMBER OF SEQUENCES: 265
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Quarles & Brady
 ; STREET: 1 South Plinckney Street
 ; CITY: Madison
 ; STATE: WI
 ; COUNTRY: US

```

      ZIP: 53701-2113
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Word Perfect 8.0
      CURRENT APPLICATION NUMBER:
      APPLICATION NUMBER: US/09/453,702B
      FILING DATE: 03-Dec-1999
      CLASSIFICATION: <Unknown>
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 60/110,955
      FILING DATE: 04-Dec-1998
      ATTORNEY/AGENT INFORMATION:
      NAME: Seay, Nicholas J.
      REGISTRATION NUMBER: 27386
      REFERENCE/DOCKET NUMBER: 960296.95017
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (608) 251-5000
      TELEFAX: (608) 251-9166
      INFORMATION FOR SEQ ID NO: 18:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 3695
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-453-702B-18

Alignment Scores:
Pred. No.: 0.000279      Length: 3695
Score: 139.50      Matches: 211
Percent Similarity: 31.32%      Conservative: 111
Best Local Similarity: 20.53%      Mismatches: 372
Query Match: 3.14%      Indels: 336
DB: 4      Gaps: 47

US-10-030-330-1 (1-843) x US-09-453-702B-18 (1-3695)
QY 63 AspAlaGluArgGlyIleThrSerIleGluGlySerProAlaIleThrValAla 82
DB 526 GACGGTCACAGCCGATTAATGCTTGAACGTAATGAGCAAAACCGCGTGGCTG 585
QY 83 AsnATGGLyAsnAsnGluGlyTyAlaLeuValAlaIleAspAspArgIleProThrIle 102
DB 586 TCTCTGGCGGAGCGGAGGCGACGATCACGGGCAATGAAGATCAATCAAGACTGAA 645
QY 103 LeuAlaIleThrSerProIleGlyArgPhe-----AspMetAspSerMetProAspAsn 119
DB 646 CTAACCTTCAACCCGCGTGAATATGTGACTCGTCCCTGAAGGCCACATAATACACAG 705
QY 120 LeuArgMetIlePleu---GluIleTyAspGluIleGlyLeu-----133
DB 706 GCAAAAGCAACACTGGGTGAGTTCACCGAACTGAAGCAGGGGTATCACTGTCTTT 765
QY 134 -----IleLeuSerGlyLyAlaGluIleLeuAsnGluIle---LeuArgThrGlu 149
DB 766 ACTACCGGAACGACAGTGTGAGGCAACAGATTAAGTTCAGTGTGATGAGCAAA 825
QY 150 GlyValProAlaGluValAlaIleLeuMetAspAsnGlyHisPheAlaAsnAspPromet 169
DB 826 ACCGTCACTGCAAGACGTGGGCGGACGATGATGAT-----GTGGCAAACTCCACCCG 879
QY 170 ArgTrpAsnGlnGlyTyTrpTrpAsnAsnLyGluPro-----LeuLeuProAsn 186
DB 880 AGCCCTAAC-----GAGCCGTCAAGTACAGGTGGTTCGTGAT 915
QY 187 GlyAsnHisAlaIleThr-----GlyCysValAlaIleThr 197
DB 916 GGTTCAGCAAGCCCTATTAAGCTTGACGTGACGTGGTACCTCCAGAGGTAAATCCGGTACAG 975
QY 198 AlaAlaIleGlnIleMetArgTyHisSer-----Trp---208

```

```

DB 976 GGAAGGCCAGCCGCTTGGCATTTGT-TCCGCAAGACACTAATGCTTAACCGTGGTGC 1034
QY 209 -----ProLeuGlnGlyGlySerPheAsp-----217
DB 1035 CATTTGGAAATTAACACAGCGGTTTACAGCCCGGGTTTCTTGACCCCGCCGGA 1094
QY 218 -----TyrHisAlaGlySerLeu-----223
DB 1095 CGTTGTTGGCTGCTTTTCAGCGAGCATATGCTGGGACATTACAACAACGCTGA 1154
QY 224 -----Val 224
DB 1155 GTTTGTTGCCGGCCGCTTGATGACGACATTCGTCATCAACCTGAAATCCTGATAAC 1214
QY 225 GlyAsnTrpSerGlyThrPheGlyIleMetTyAspTrpIleAsnMetProGlyAsnPro 244
DB 1215 GGTGGTTGGGGGACAGTTACGCG-----AATCTGACGGTAAAGA 1256
QY 245 AspLeu-AspAsn---LeuThrGlnSerGlnValAspAlaIleThrIleMetArgAs 263
DB 1257 TGCTATATGACAACCTGTGACACAGCCTCAGCCCGGAACCCGCTATAGCGGTGCCG 1316
QY 263 PValSerAlaSerValSerMetSerPheTyGluAsnGlySerGlyTyThrValTy 283
DB 1317 TGCTGAAGCTTCTACGGCATCGGCGCTGGCAAAATATGCTGATGAGTGAAGTGGCA 1376
QY 283 rVal-ValGlyAlaLeuArgAsn---AsnPheArgTyAspArgSer-----297
DB 1377 GATTACTCTCGGCTCTACGGCGGGGTGAATTAAGATTAATGCCAAGCTTAATGACAGAA 1436
QY 298 -----LeuGlnLeuHis-----ValArgAlaLeuTyThrSerG 309
DB 1437 TCGCGACGAATATCGGCAAAAGTAAACCGTGGTGTGATGCTTATCTCAACACAGTC 1496
QY 309 GlnGlu-----311
DB 1497 GAAAGTCTCTGTGCGACAAATCAGTAAACCGCGGCAACACACCGTGAACGCTGTG 1556
QY 311 rPHisAspMetIleArgGlyGluLeuAlaSerGlyArgProValTyTyAlaGlyAsnA 331
DB 1557 GCGCAAAAGATGCG-CATGAGACGCTATCAGTGT-----1590
QY 331 snGlnSerIleGlyHisAlaPheValCysAspIleTyAlaSerAspGlyThrPheHisP 351
DB 1591 -----CTTGGGTTGTGCGCAAGTTTGAACGGGGACCCGCTTGAAGGGGAGCCGTTT 1642
QY 351 heAsnTrpGlyTyTrpGlyGlyValSerAsnGlyPheTyTrpLeuThrLeuSerProT 371
DB 1643 CCAAGTTGACCGCAAAAGGT-----AACGGTTCTATGTGCTACGTTGACTACAGGTG 1696
QY 371 hrSerLeuGly-----374
DB 1697 GAAAGCGGGCGAGCTTCCGTCACGCTCTCTTCAACGCCGACACAGCACCCAGAAAG 1756
QY 375 -----IleGlyGlyGlu-----378
DB 1757 CCGCGCAGTTGACGCTCATCGCGGAGAGATTCATCAAGCAACTCTACGCTTGTGGCG 1816
QY 379 -----GlyIleGlyPheThrIle-----385
DB 1817 ACAATTAAGCTTCGACCGTCAAAACGACGAGCACTACAGTGTTAAGCGGTGCCAGCA 1936
QY 385 yrcGlnGluIleThrGlyIleGluPro-----AlaArgT 397
DB 1877 ACGGGAACCGGTTCACCGGCTGAAAGCCAGATGACCAAGTGTTAAGCGGTGCCAGCA 1936
QY 397 hrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluIle 417
DB 1937 CGGGAGTGAAGCGTCTTCA-----GACG 1960
QY 417 LuTyIleSerGlnSerGlyLeuAsnValGlyTyTrpSerIleTyAsnThrGlyGluG 437

```

Db 1961 GAAACTGACAGAGAAAGTAAATGGGGTCTACGTCTGACCTTAACGCTGGATCGCG 2020
QY 437 InsertAsnLeuAspLeuGlyTyrArgLeuAsn----- 447
Db 2021 CGGGTCAGTGTCTGTGATGCCGCGAGTACAGCGCAAAATGCCGTGTCTGACCGCATCG 2080
QY 448 -----LysAlaAspGlyGluValIleGluValIleThrSerSerIleAsnIleG 464
Db 2081 TGCCTGAACCTTCAGCTGACGCATCTTAAGCTGAGTTGCTGATATGACAGTAAAGTTA 2140
QY 464 eTTPtGlyTyrGlyGluHisProGluSerPheSerLeuAlaProAsnGlnLeuSerG 484
Db 2141 Ar-----LleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGluGlnTTPG 502
QY 484 InGly-----LleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGluGlnTTPG 502
Db 2156 ATGACAGCTGCTGCTAACCAGATACCCG----- 2184
QY 502 LUProValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsnThrThrAsp 522
Db 2185 --ACCGTTGTGACACACCTATGGTAACCCGTGCGAGGCGAGAGATTACGCTGACTTAC 2242
QY 522 roAsnAsnValVal-----ValThrValAspAsnAsnGlyLysLeuS 537
Db 2243 CCCAGGGTGTGACCAAGCAAGCGGGGAATACAGTAACTAATGCGCGAGTAAAGCGG 2302
QY 537 eTTPeValProAsnSerPheValAlaAspLeuAsnSerTyrGluHisSer----- 553
Db 2303 ACATTGAGCTTATGTCACAGGTGCGGAGAACACAAATATTTCCGCTGCGTGAATGCTG 2362
QY 554 -----ThrIleThrValGlnPheAsnSerAsp---SerProAspGluIleArgThrP 570
Db 2363 CTCAGAGACGCTGACACCGTGAATTCACGCGCATCCAGCACCGGTCAGGCAAACTGCG 2422
QY 570 roValAlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTTPY 590
Db 2423 AGCTGAGACGCGCTGCTCAAAAAGTGGCAACGCGCAAGAGCTTACGCTGACGCGCA 2482
QY 590 aIMetAlaGlu-----ValProGlyGly-----SerSerAspTyrPro- 602
Db 2483 ACCTTGAGAGTAAATAATGTTAACCTGTTCCAGGGAGCCTGTGACCTTAATCTGCCCG 2542
QY 603 -----ValValTyrSerLys----- 607
Db 2543 GGGGTGTCAAGCGCTTACAGCGGAAATGCTCGGTGAAGAACCAACGATGAGGGGAAG 2602
QY 608 -----AspValLeuThrLeuSerGluGlyAspTyrThrLeuTTPYrArgPheSerI 625
Db 2603 CAGAGTTGACAGGTGTTTTCAGTACTGCCGGAACGTAATGATCAGCGCATCGCAGCG- 2661
QY 625 IeAsnAsnGlnLysAspGluTTPYrLysLysIleGlySerValSerValLysThrProThG 645
Db 2662 --AATAGCCAGCCTTCGAATACGAGACTAATACGTTTGTAGCCGTAAGGCTACCGCAA 2719
QY 645 IuTyrThr-----HisProLeuPheGluValIleGlnHisAsnGlnT 658
Db 2720 CGGTCTCCGATTTAGATGATGATGCAATATGACACTGGCGGCAAGCATGCCAAACAG 2779
QY 658 hSerThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAsp-----PheThrL 676
Db 2780 CGTATTAAGTATGAGGTGATGCAATTAACCACTGTTGAAAGATAGCAGAGTACCG 2839
QY 676 eu-----LysAsnLeuGlyLeuProPheAsnGlyGluLeuValValAlaPhea 692
Db 2840 TCACTGCCAGCCCGCAATTTAGTTCTACTGCCCAATGGGAGC-----G 2884
QY 692 rGlnThrGlnSerSerGlySerLeuTTPAlaIleGlnGluThrValHisIleLysG 712
Db 2885 CGAAACTAATAGACAGACAGCGATTTTTCACGCCACAGACACTGTGCGCGCAAA- 2943
QY 712 InGlyGluThrPheValTyrLysProValIleGlnGlyProIleProAspGly-----S 730
Db 2944 -----TATACACTCAGCGCAAAAGTGAATGATGAGCGGCAAGCTACAGAAAT 2986

QY 730 eTTPArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnGlnLeuTyrLeuLysGlyL 750
Db 2987 CGAGCAAAATCGCCGAATCTTAATTCGTGCGGATGATACAAATGCACTACTACCGCAT 3046
QY 750 ysArgAsnTyrThrValLysIleValAsnGlyThrAlaValIleGluIleSerSerG 770
Db 3047 CATCTGATGATGATTTCTCTGCTGCGGATGCGGATGATGATGATGATGATGATGATG 3106
QY 770 InGluIleArgValIlePheProAsnProAlaArgAspTyrValIleGluIleSerAlaPro- 788
Db 3107 TGAATGCGCAATTAACCCGTTGGGGGAATATGTCGTCGATTAAGACCGCAAG 3166
QY 789 -----CysIleProGlnGluThrSerIleIleLeuPheAspL 801
Db 3167 GGGTGAAGCAAGATTAATCACTGCTGCGGCAAAATGACCAATTC-----G 3217
QY 801 euserGlyLysIleValIleMetLysAsnSerLeuSer---AlaGlyHisGlyArgMetAspV 820
Db 3218 TGAGCGCAAAATCAGCGCGTACATTTAGTACCAAGCAAGCCTGGTCTATACGTTCAAT 3277
QY 820 aIserArgLeuProAsnGlyAlaTyrIleLeuLys-----Y 832
Db 3278 TTAACCCCTGACGTATGCGCGGTACGAAATGAAGCCAGTGCAGCATTAACCGCG 3337
QY 832 aIAspGlyTyrThrThrLys 838
Db 3338 TGGATCCGATACGCAAG 3357

RESULT 18
US-08-702-572-14
; Sequence 14, Application US/08702572
; Patent No. 5965386
; GENERAL INFORMATION:
; APPLICANT: Kerry-Williams, Sean M
; APPLICANT: Gilbert, Sarah C
; TITLE OF INVENTION: Yeast Strains and Modified Albumins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Centeon L.L.C.
; STREET: 1020 First Avenue
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-1310
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702,572
; FILING DATE: 11-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NO 95/23857
; FILING DATE: 1-MAR-1995
; APPLICATION NUMBER: GB 9404270.2
; FILING DATE: 5-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Naomi Biswas
; REGISTRATION NUMBER: 38,384
; REFERENCE/DOCKET NUMBER: CE0114 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610/878/4294
; TELEFAX: 610/878/4221
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4106 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

Db 3303 GATCGTATTCGATCAACCAACTCTTAATAAACTTCTCTCCGTAGGACCAACCATGAT 3362
QY 737 Alaphavalasn-----Glyginginglnleu-----Tyr 746
Db 3363 TATTTTAAACATATTTTGTGATGGCCACATTTTGTGTATACCTCATGTTT 3422
QY 747 LeuLysGlyLysArgsnTyrThrValLysIleValasnGlyThrAlaValaIle 766
Db 3423 ATGAAATCAAGGAAAGG-----ATCAGAGGTCAAGCGGAAACGTA 3467
QY 767 Glu-----SerSerGluIleArgValPheProAsnProAlaArg 780
Db 3468 GAATTCGATATCATGATACAGACTGTGATGATTCATCTTGACATATGAACTTCC 3527
QY 781 AspTyrValGluIleSerAlaProCysIleProGluThrSerIleIleuPheAsp 800
Db 3528 GGAATTAATCTAG-----CCCGAAGAGTTCGAGCACTTCGATTTTGTAT 3569
801 LeuSerGlyLysIleValmetLysAsnSerLeuSerAlaGlyHisGly 816
3570 TTGTCCGATGAAACACATCTTGCAGCTTGTCTTCATCAAAAACGCT 3617
RESULT 19
5521093-4
Patent No. 5521093
APPLICANT: LEMOINE, YVES; NGUYEN, MARTINE; ACHSTER, TILMAN
TITLE OF INVENTION: YEAST VECTOR CODING FOR HETEROLOGOUS
GENE FUSIONS LINKED VIA KEX2 CLEAVAGE SITE AND CODING FOR
TRUNCATED KEX2 GENES
NUMBER OF SEQUENCES: 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393, 025
FILING DATE: 23-FEB-1995
Prior APPLICATION DATA:
APPLICATION NUMBER: 191, 354
FILING DATE: 07-FEB-1994
APPLICATION NUMBER: 26, 121
FILING DATE: 04-MAR-1993
APPLICATION NUMBER: 500, 885
FILING DATE: 29-MAR-1990
SEQ ID NO: 4
LENGTH: 4732
5521093-4
Alignment Scores:
Pred. No.: 0.000418 Length: 4732
Score: 139.50 Matches: 167
Percent Similarity: 32.67% Conservative: 80
Local Similarity: 22.09% Mismatches: 278
Match: 3.14% Indels: 231
Gaps: 43
US-10-030-330-1 (1-843) x 5521093-4 (1-4732)
QY 166 AsnAspProMetArgTyrAsnGlnGlyTyrProTyrAsnAsnLysGluProLeuPro 185
Db 1748 AATGATCCGCTTTTGAGAGGAG-----TGCACCTTGCTCATCAATTTTCC 1798
QY 186 Asn-----GlyAsnHisAlaTyrThrGlyCysVal 195
Db 1799 GGCAGTATATAATGCTTCTGATCTGTGTACAAATAATATACAGCGCGAGGCGCTG 1858
QY 196 AlaThrAlaAlaGlnIleMetArgTyrHisSerTyrProLeuGln----- 211
Db 1859 GCGCCATTTGATGATGCGCTTGACTACGAAAGAAAGAGACTTGAAGATAATTTTGC 1918
QY 212 GlyGluGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGlyThrPhe 231
Db 1919 GCGAAGGCTTCTGGATTTCAAC----- 1942
QY 232 GlyGluMetTyrAspTyrPheAsnMetProGlyAsnProAspLeuAspAsnLeuThrGln 251
Db 1943 -----GACAAATACCAATTTACCT-----AAACCAAGATTATCTGAT----- 1978

QY 252 SerGlnValAspAlaTyrAlaThr-----LeuMetArgAspValSerAlaSerLeuMet 270
Db 1979 -----GACATACATGATGATGAGATGAGTGCAGGTGAATTAACCTGCAAAAAAGGTAC 2029
QY 271 SerPheTyrGluAsnGlySerGlyThrTyrSerValTyrValGlyAlaLeuArgAsn 290
Db 2030 AATTTTGGCGGTGCGGTAGGT-----TACACGCTAAATCTCAGGCATA----- 2077
QY 291 AsnPheArgTyrLysArgSerLeuGln----- 299
Db 2078 -----AGAACTTATCCCGGTATATCATCTACGGAAGATGAAGCTGCGCTC 2122
QY 300 -----LeuHisValArgAlaLeuTyrThrSerGlnIleTyr----- 311
Db 2123 TTGATTTATGCTCTAGACGTAAACGATATATAT-----TCAAGCTCATGGGTCCCGCTGAT 2179
QY 312 -----HisAspMetIleArgGlyGluLeuAlaSerGlyArg 323
Db 2180 GACGGAAGACATTTACAGGCCCTAGTACGCTGTGAAAAGGCTTAGTAAGTAAGT--- 2236
QY 324 ProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyr 343
Db 2237 -----GTTACTGAGGAGAGATTCGCAAGGAGCGATTAACGTT-----TTT 2278
QY 344 AlaSer-----AspGlyThrPheHisPheAsnTyrProGlyGlyValSerAsnGly 361
Db 2279 GCCAGTGGAAAGTGGTGAAGCTGCTGTGATTAATTCAGACGGCTTATCAATATCC 2338
QY 362 PheTyrLysLeuThr-----LeuLeuSerProThrSerLeu 373
Db 2339 ATATATCTATCTACTATTGGCGCTATGATCACAAGATTCATCTCTTATTCGGA 2398
QY 374 GlyIleGly-----GlyGluGlyIleGlyPheThrIleTyrGln 386
Db 2399 GGTGTGTCGCCCTCATGCGAGTCAGCTATTCAGGTTCAGCGCAATATATCAATTCG 2458
QY 387 GluIleIleThrGlyIle-----GluProAlaLysThrProAla 399
Db 2459 AGTGAATTCACAGCGAGATGACAGTCAATATGCCACGCGGAAAGCTGCGGCTCCATTA 2518
QY 400 GluAlaGlyThr-----AspAlaLeuProIleLeuAlaLeuLysAspIle 414
Db 2519 GCGTCCGGTGTATACACTTTGTTACTAGAACCAACCAACCTGATGGAGAGACTA 2578
QY 415 GluAlaGluTyrLysSerGluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGly 434
Db 2579 CAGTAT-----TTATCAATCTGTCTGCGGTA 2605
QY 435 GluGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluValIle 454
Db 2606 GGGTGAAGAAAGACCGTGCAGGAGATTTGGAGA-----GATAGCCCATGAGGAATAATAC 2662
QY 455 GluValLysThrSerSerIleAsnIleSerTyrGlyGlyGlnHisProGluSer 474
Db 2663 TCTCATCGC-----TATGCGTTTGTAAATCATGCCAT 2698
QY 475 PheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuTyrArg 494
Db 2699 AACTTAATGAAATGTCACAGACCTGGAGAAATGTTACACACAAACCTGTTTACCTG 2758
QY 495 ArgThrGlyThrGluGlnTyrProValArgHisIleGlnGlyGlyTyrValAsnSer 514
Db 2759 CCACACA-----TTGTAATGTTTCCAG 2779
QY 515 IleLysValAsnThrThrAspProAsnAsnValValThrValAspAsnAsnGly 534
Db 2780 TCCAAACATCCACAGGAGAGACATTAATGCTGCTATACATA-----TCAGAAAAA 2833
QY 535 LysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThr 554
Db 2834 AGCTTCA-----GATGCTAACTTCAAGAAATAGACAGCTGACG 2875

Db 1405 AGGAGTGAATGAGGGGTAGCATGTGGATATTATGAAAGATGATGACATACGACAGAAC 1464
 Qy 268 -----ValserMetSerPhe-----TyrGluasnGlySer 277
 Db 1465 CATGTGACAGACCTTAATGAAGCATGAGCTCTGGTGAAGAAAGCTCCATCTAGATGAGTACT 1524
 Qy 278 GlyThrSerValTyrValValGlyAla-----LeuArgasn 290
 Db 1525 GATCCAGAAAGCCTTCTCTGGATGA-GCTTACCGAAAGTACGGTGTGTATTCGTAAT 1583
 Qy 291 AsnPharGlyTyrLysArgSer----- 297
 Db 1584 AGCTGACGAAATGAAAGTCTGTCGATTAATCATGTGGAAGCTCGTGTGCAACAA 1643
 Qy 298 -----LeuGlnLeuHisValArg-AlaLeuTyrThrSerGln 310
 Db 1644 GGCATTAACCTGTGGAGCTGCTGACAGTCCCATTAACGTGGAGTTATGTTTCCCA 1701
 Qy 310 uTrpHisAspMetIleArgGlyLubLeuAlaSerGlyArgProValTyrTyr 327
 Db 1702 -----AlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyrAl 344
 Qy 328 -----AlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyrAl 344
 Db 1734 TTAATCCCGCCGCTTACACAAACCTTAGAATAGCATTC-TTCTCAAGCAGAGGGCCGAG 1790
 Qy 344 AsnAspGlyThrPheHisPheAsnTrp-----GlyTyrGlyValSerAsn 361
 Db 1791 AATAGATGTTGAATTAACCCCAATGATGCTGCTCCAGGTTACGGA----- 1836
 Qy 361 yPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyValGlyLysIleG 381
 Db 1837 -----ATTATCCATCCCGCCGAGTGTGATGGCGGA----- 1869
 Qy 381 yPheThrIleTyrGlnGluIleIleThrGlyIleGluProAlaLysThrProAlaGluAl 401
 Db 1870 -----GCTGACTTCATGTCGTGACACTGTCGATGCT-ACCTCCACATGTCAG 1913
 Qy 401 eGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluAlaGluTyrLysSer 420
 Db 1914 CGGTGTCGTGCTCCTCTCTTAAGCGGGGCAAGGCCGAGGAATATCTACAATCCAGA 1973
 Qy 421 -----GluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrG 434
 Db 1974 TATATTAAGAAAGCTTCTGAGAGCGGTGCAACCTGCTGAGGAGATCCATATCTGG 2033
 Qy 434 yGluGluGlnSerAsnLeuAspLeuGlyTyrArg----- 445
 Db 2034 GCAGAGTACACTGACCTTGACCAAGGTGCTTGTACCTTACCAAGTCTGGGA 2093
 Qy 446 -----LeuAsnLysAlaAspGlyLysValIleGluValLysThrSerSerIleAsnIle 464
 Db 2094 AATCTTAAGGCTATTAAGCGCACACCTCCCAATTTGTTGATCTGGGCAACAGTCT 2153
 Qy 464 rTyrTyrGlyTyrGlyGlnHisProGluSerPheSerIleuAlaProAsnGlnLeuSerG 484
 Db 2154 CTACACGACCTTGGCGAGTAC-----TT 2177
 Qy 484 nGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThGlu-----GlnTr 501
 Db 2178 GGGTGTGAGCTTATAGAGTCTCTACGCAAGAACTATACCTGACATCTGCTGAGT 2237
 Qy 501 pGluProValArgHisAlaGlnGlnGlyTyrValAsnSerIleLysValAsnThrHis 521
 Db 2238 GCAC-----ATTAGTACGTAGGGGACACGAGTACAGAACTTTGAGATCTATGCAACTGA 2294
 Qy 521 pPro-----AsnAsnValValValThrValAspAsnAsnGlnGly 535
 Db 2295 GCCATGATTAACCTTTTCTCAGTGAAGTGTATTTCTAGACAACATCCAGATTGT 2354
 Qy 535 sLeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrI 555
 Db 2355 CCTTAGGGTG-----AAATATGATGTAGAGGCTCTTGAGCCAGGTCTCTA 2399

Qy 555 eThrValGlnPheAsnSerPheProAspGluIleArgThrProValAlaPheAla 575
 Db 2400 TGTGGAAAGATTAATCTATGATCCA-----ACAACGCCAGTT----- 2439
 Qy 575 uSerThrGlyAlaThrAlaAspValIleSerLeuGlyTyrValMetAlaGluValPr 595
 Db 2440 -----ATTGAACGACGATCTTG----- 2457
 Qy 595 oGlyGlySerSerAsnTyrProValValTyrPheLysAspValLeuThrLeuSerGlu 615
 Db 2458 -----AACCAATTTGTATTTCCGAGAG-----TTCACTCCTGGAA 2495
 Qy 615 yAspTyrThrLeu-----TyrTyrArgPheSerIleAsnAsnGlnLysAspIleTyrLys 634
 Db 2496 CAATTACCCCTCACTGAT----- 2517
 Qy 634 sIleGlySerValSerValLysThrProThrGluTyrThrHisProLeuPheGluVal 654
 Db 2518 -----GATATTAATGATCCAGAAATGTCACACACACTTCTCACTGTGCC 2564
 Qy 654 yHisAsnGlnThrSerThrTyrThrLeuAspMetAlaHisAsnArgValLeu-----Pr 672
 Db 2565 TGAGGAGTGGACGTTCTCTACCGATGACACATACCTGGGACTACGCTGTACAGACC 2624
 Qy 672 oAsp-----PheThrLeuLysAsnLeu-----G 680
 Db 2625 AGATGCAATTTGTGTCCCATACACAGCTAGATATCTCCGCTCAGCTCAATCC 2664
 Qy 680 yLeuProPheAsnGlyLysLeuValValVal-----PheArgGlnThrGlnSer 697
 Db 2685 AATGCCGGAACCTGGAGCTAGATGACTGATTAATCTTACCTCCACCTCATAGTCC 2744
 Qy 697 r-----SerGlySerLeuThrAlaGln 706
 Db 2745 GGGCTTCCTTGTAGGATTAACGAGTAGAATAACTCAAGCCTTGGTATTAACAG 2804
 Qy 706 uThr-----ValHisIleLysGlnGlyLysThrPheValTyr 719
 Db 2805 GACATACCTGACACTACACTGATCTCAATTTGAATTTCAATTTCTAATCATCTATGC 2864
 Qy 719 sProValValGlnGly-----ProIleProAspGlySerTyrArgAlaThrLeuHisAl 737
 Db 2865 CCCAATTAAGCAACTATTAATGCTGCTGGAACCTGAAATCGAGCGTTGAAG 2924
 Qy 737 apheValAsnGlyGlnGlnLeuTyrLeuLysGlyLysArgAsnTyrThrValLysI 757
 Db 2925 CGTT-----GGTATGAGAGAGTCTTCATAAAGGC----- 2955
 Qy 757 eValAsnGlyThrAlaValAlaIleGluSerSerGluGlnIleArgValPheProAs 777
 Db 2956 -----ATTGAAGTCTCTGAAGGCACCGCAGAGTTGAG----- 2988
 Qy 777 nProAlaArgAspTyrValGluIleSerAlaProCysIlePro-----GlnIleThrSerI 796
 Db 2989 -----ATTAGATATGGCAACCAAGTTCGCAATTCAGATCTAGACT 3032
 Qy 796 eIleLeuPheAspLeuSerGlyLysIleValMetLysAsnSerLeuSerAlaGlyHisG 816
 Db 3033 GTACCTTATAGACAGTAAGAGCAATTAAGAGGCTTATAGAGGAACCAACAGCAGAGA 3092
 Qy 816 yArgMetAspValSerArgLeuProAsnGlyAlaTyrIleLeuLysValAspGlyTyrTh 836
 Db 3093 AGAGGTGTAGTGTAGTATCCTAAGCCTGAGATTATCAATAGTACATGATGTTACAG 3152
 Qy 836 rThrLys 838
 Db 3153 GGTGAG 3159

Search completed: June 2, 2003, 01:07:31
 Job time : 375 secs

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

OY		646 rThr 647 	
Db	1574 GACG 1577		
RESULT 2	BH770729/c		
LOCUS	BH770729	1557 bp	DNA linear GSS 01-MAY-2002
DEFINITION	LMGtag478 MG1363 Random Sequence Tag Library Lactococcus lactis		
ACCESSION	BH770729		
VERSION	BH770729.1		
KEYWORDS	GSS.		
SOURCE	Lactococcus lactis subsp. cremoris.		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
REFERENCE	Lactococcus.		
AUTHORS	1 (bases 1 to 1557)		
TITLE	Boletín, A., Ehrlich, S.D. and Sorokin, A.		
JOURNAL	Studies of genomes of dairy bacteria Lactococcus lactis		
COMMENTARY	Sci. Aliment. (2002) In press		
	Contact: Sorokin A		
	Genetique Microbiome		
	INRA		
	CRI INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France		
	Tel: 33 1 34 65 25 16		
	Fax: 33 1 34 65 25 21		
	Email: sorokine@jouy.inra.fr		
	best homologue in strain IL1403 is ywfg (76%)		
	Class: shotgun		
	High quality sequence start: 30		
	High quality sequence stop: 1527.		
FEATURES	Location/Qualifiers		
source	1..1557		
	/organism="Lactococcus lactis subsp. cremoris"		
	/strain="MG1363"		
	/db_xref="taxon:1359"		
	/clone_lib="MG1363 Random Sequence Tag Library"		
	/note="Vector: pSGM2; Site 1: SmaI; library of chromosomal fragments of L.lactis strain MG1363 was prepared by partial AluI digestion or by sonication."		
BASE COUNT	403 a 318 c 253 g 382 t 1 others		
ORIGIN			
Alignment Scores:			
Pred. No.:	0.151	Length:	1557
Score:	114.00	Matches:	106
Percent Similarity:	32.58%	Conservative:	68
Local Similarity:	19.85%	Mismatches:	190
Identity Match:	2.57%	Indels:	170
	17	Gaps:	24
US-10-030-330-1 (1-843) x BH770729 (1-1557)			
OY	256 AlaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTyGluAsn	275	
Db	1499 GCCTAACAGCAGACAGCAAAAATAATTCACGATATACTATTCAATCA-----TCAAA	1446	
OY	276 GlySerGlyThrTyrSerValTyrValValGlyAlaLeuAlaGlnAsnAsnPheArgTyrIys	295	
Db	1445 GGACCACGCACAAGCAGGAAGTAACGGTTATGATGTCAAACTGTAAATTAATGTTTATCA	1386	
OY	296 ArgSerLeuGlnLeuHisValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIle	315	
Db	1385 AGAAT---CAAAGACATGTGATGTAACCTACATGATGATCGACACAAGGCCAAGTTTTC	1329	
OY	316 ---ArgGlyGluLeuAlaSerGlyArg-----ProValTyrTyrAla-----	328	
Db	1328 ACAGAAGAAAGATTGTCGTGGAACATGATGACCTCGCATTTATACAAAGTGCTGAT	1269	
OY	329 ---GlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyrAlaSerAspGly	347	
Db	1268 ATTAAATTTTACACGAGTAAAGATATGATATTGATGATTAATCCGATCCAGATCAAGGA	1209	

[illegible]

Oy 603 Valva1TtpserlyaspvalleuthrleusergluGlyAspTyrThrleutryrArg 622
 Db 230 ACCGTT-----GATACGATTAATTAACACAGATAGGTTACGAATG----- 186
 Oy 623 Pheserileasnasnclnlyaspclutrplys----- 633
 Db 185 GTTTCAGGATTAATCCAGAGATGAACTAAATTCGCGATGACCTCAACATTATGTT 126
 Oy 634 -----Lysilegylseral-----SerallysthrProthglutyrThr 647
 Db 125 GTGAGATTAACAGGTACACTACTACTGAACTGAAATTAATCTGTTAATGAATGATCATT 66
 Oy 648 H1SPoleuphegluValGlyHisasnclntrSertrrtyr 661
 Db 65 CATTATGTTTATGATTAATGAGATTAAGGACAGCTGACACCTAT 24
 RESULT 3
 AK018613
 ACUS
 ACCESSION AK018613 2178 bp mRNA linear HTC 19-JAN-2002
 VERSION Mus musculus adult male cecum cDNA, RIKEN full-length enriched
 KEYWORDS full insert sequence.
 SOURCE AK018613.1 GI:12858408
 HTG: CAP Trapper.
 Mus musculus (strain:C57BL/6J) adult male cecum cDNA to mRNA,
 clone:J130012D09.
 Mus musculus
 ORGANISM Mus musculus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE
 AUTHORS 2
 TITLE Carninci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499377
 PUBMED 11042159
 REFERENCE
 AUTHORS 3
 TITLE Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, K.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE
 AUTHORS 4
 TITLE Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Aizawa, K., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
 Saito, T., Okazaki, Y., Goto, J., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleischmann, W., Gaasterland, T., Gissi, C., Kling, B., Kochiwa, H.,
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
 Quackenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M.,
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
 Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
 Carninci, P., de Bonaldo, M.F., Brownstein, M.D., Bult, C.,
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
 Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Sessa, T., Shibata, Y., Storch, K.F., Suzuki, H.,
 Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,
 Wyshaw-Borls, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.,
 and Hayashizaki, Y.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409 (6821), 685-690 (2001)
 MEDLINE 21085660
 PUBMED 11217851
 REFERENCE
 AUTHORS 5 (bases 1 to 2178)
 TITLE Aizawa, K., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
 Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,
 Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hasegawa, T.,
 Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Hori, F.,
 Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Koyama, S.,
 Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koyama, S.,
 Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
 Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
 Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
 Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
 Tanaka, T., Tejima, Y., Toy, T., Yamamura, T., Yamana, I.,
 Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
 Hayashizaki, Y.
 COMMENT Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in RIKEN
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in RIKEN contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5' GAGACGAGATTCGACGATTAATTAATCCCGCCCCCCC 3']. cDNA was
 prepared by using trihalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. cDNA went
 through one round of normalization to Rot = 10.0 and subtraction to
 Rot = 185.2. Second strand cDNA was prepared with the primer
 GAGACGAGATTCGACGATTAATTAATCCCGCCCCCCC 3'. cDNA was cleaved
 with BamHI and XhoI. Vector: a modified pluescript KS(+) after
 bulk excision from lambdaBamBla I. Cloning sites, 5' end: SalI, 3'
 end: BamHI. Host: DH10B.
 Location/Qualifiers
 1..2178
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="RANTOM:DB:9130012D09"
 /db_xref="MGI:1904161"
 /db_xref="taxon:10090"
 /clone="9130012D09"
 /sex="male"
 /tissue-type="cecum"
 /clone.lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 74..1807
 /note="Immunoglobulin domain containing protein
 data source: Pfam, source key: PF00047, evidence: ISS
 putative"
 /codon_start=1
 /protein_id="BA31307.1"
 /db_xref="GI:12858409"
 /db_xref="MGI:1918851"
 /translation="MELAGFCCSWVILALPTTPMPAPAAHPIHRAILSTNOM
 ERPLALATLVNDSVTEQREMAVYGCNTNDNINIHVNSLLIVLERRKLSND
 KTLIIIVREDSDSYICEYQGEFVORNSLTATYPPDPVSKIKDSGVAAGVVE
 VMGNTVNFVEAQSPPVATYATYLPDSITPTGTPTDAVSRHEGMYRCLVSNP

VTNLSRLGVKVOLEKVTAPNIEPTLALVENATSVLTCTKSHORGVHVELKGP
LRSPRLTLEULEUFTYRARGAARGHRLGTHRGUINGINTPGLUProValArgHisala 507
TQPRASVSTTEAMNLSITLYCRADSDIGPECEVNMWGSOASBPJKTINTXPDQVEI
HOGITCSNDVGLASASVLYVNVLOSMSVGAAGVIGLIVLALALALIGL
FLYSTKDMTRRSASDTSNTIPETVMSOSTPESRDNKPTVYDNPKEGEGARG
KMSLIPDSPEDEFEKKPPSAPEGRPRKPLQIPKQPLMPPGGNESNVEKLNSN
HSLVCKITPPSA*

BASE COUNT 541 a 631 c 529 g 477 t
ORIGIN

Alignment Scores:

Pred. No.: 0.642 Length: 2178
Score: 111.00 Matches: 79
Percent Similarity: 37.54% Conservative: 55
Best Local Similarity: 22.13% Mismatches: 135
Query Match: 2.308 Indels: 88
DB: 11 Gaps: 21

-030-330-1 (1-843) x AK018613 (1-2178)

488 ThrileThleuleuTyRArgAArgHRLGTHRGUINGINTPGLUProValArgHisala 507
||||| : : : : : ||| ||||| |||
230 ACCCTCACAC---GTCATCAGACGACCTGTACAGAGAGAGGAAATGGCAGCTTTCTAC 286
508 GlnGlyGlyTyRValAsnSerIleTySValAsnThrThrAspProAsnAsnValVal 527
||||| : : : : : ||| ||||| |||
287 TGTAACTANTGCGATACATACATCCATCGGCTGCCAACACCTCTGCTTG 346
528 -----ThyValAspAsnAsnGlnGlyLysLeuSerIleValPro 540
||||| : : : : : ||| ||||| |||
347 TTGAATGAGCCATGACCTGTCTGATACATACAGACCTCACCATCTCTATGTCAG 406
541 AsnSerPheValAlaAspLeuAsnSerTyR-----GlnHis----- 552
407 CGGAG-----GACTCGGGGCTTACTGCTGTGAGAGTCAGCATGGCTTGAGGTT 457
553 -----SerThrIleThrValGlnPheAsnSerAspSerProaspLuile 567
458 CAGAGAACTAACACCTCGCTGCGTACGATGATGCTCGAT----- 502
568 ArgThrProValAlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerIeu 587
||||| : : : : : ||| ||||| |||
503 -----CCTGTATCATACAGCTGACCTGACGAGTACGCTGGGAGTGGTGAGG 556
588 -----GlyTrpValMetAlaGluValProGlyGlySerSerAsnTyProVal--- 603
557 ATGCAAGCAACACCTGACACTTACAGGCTGAGAGACAGCTCTCCCTGCTGCTAT 616
604 ValTrp-----SerLysAspValLeuThrLeuSerGlnGlyAspTyRThreU--- 619
617 GCGTGTATCTCCCGACGAGCTATCCAGCCCTTACACAGAAATTCATCATTCAT 676
620 -----TrpTyRArgPheSerIleAsnAsnGlnLysAspGlu 631
677 GCTGTGTCCAGGACATGAGGCGCATGTACAGGCTGCTGCTCACCCTGTCAACAAC 736
632 TrpLysIleGlySerValSerValLysThrProThnGlnTyRThnHisProleuPhe 651
||||| : : : : : ||| ||||| |||
737 TTGTCCTGCTAGGTGTGCTCAAGTCCAGTCTTGAAGAGTACCTGACCCAGAAATATC 796
652 Glu-----ValGlyHisAsnGlnThrSer-----ThyTyRThreLysPmet 665
797 GAGTTCCTCCACCTGGCCCTGTGTGAGAAATGCCACCTCTGTACACTGACCTGCAAAAC 856
666 AlaHisAsnArgValLeuProAspPheThreLys----- 677
857 AGCCACCAAGAGGTGTGCTCATGCTTCTAAAGGTCAACCACTGAGCCCTAGTATG 916
678 AsnLeuGlyLeuProPheAsnGlyLysLeuValValPheArgGlnThrInserSer 697
||||| : : : : : ||| ||||| |||
917 GCGCTGACACTGTCTATCCCAAAACAGAACCTGACATCCATGAGCTTCAGAGATGAT 976
698 SerGlySerLeu-----TrpAlaAlaGln-----GluThrValHisIle 710

DB 977 ATGGGCGCTATGAGCTGTGAGCTGTGAGACTGGGGTAGCCAGCAGCGCTCCCTG 1036
||| : : : : : ||| ||||| |||
QY 711 LysGlnGlyGluThrPheValTyRysPro-----ValValGlnGlyProIle 726
||| : : : : : ||| ||||| |||
DB 1037 AAGCTC-----ACATCAACATATGGCCCCGACCAAGTAGGAATATCACCAGGGCCAGG 1090
QY 727 ProAspGlySerTyRArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnGlnLeuTyR 746
||| : : : : : ||| ||||| |||
DB 1091 -----TCTGAGGTGTGCTGACTATTTGAGCAATGCTCACTCCAGCTTGACCTGTAT 1144
QY 747 -----LeuysGlyLysArg-----AsnTyRThrValLys 756
||| : : : : : ||| ||||| |||
DB 1145 TCGCCGGCGAGCTCCATACACAGCTGCTGAGCTATCACTGACCATATGACATCTTAAAG 1204
QY 757 ILeValAsnGly-----ThraValGlnAlaIleGlySerSerGlnGlu 771
||||| : : : : : ||| ||||| |||
DB 1205 GTGCTTGTATGGGAGCAGCTGAGCATCGAAGCACTGAGCAGGAGCATCAG 1255

RESULT 4

BM468550/c 1093 bp mRNA linear EST 05-FEB-2002
LOCUS
DEFINITION AGENCOURT_6432286 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5535892
5', mRNA sequence.
ACCESSION BM468550
VERSION BM468550.1 GI:16517592
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 1093)
AUTHORS NIH-MGC http://mgi.mcl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabrs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM12224 row: 1 column: 05
High quality sequence stop: 644.
Location/Qualifiers
1. 1093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5535892"
/clone_1lb="NIH_MGC_71"
/tissue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site 1: Noti;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."
BASE COUNT 254 a 304 c 257 g 277 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 1.15 Length: 1093
Score: 104.00 Matches: 79
Percent Similarity: 33.14% Conservative: 36
Best Local Similarity: 22.77% Mismatches: 120
Query Match: 2.34% Indels: 113
DB: 13 Gaps: 19

US-10-030-330-1 (1-843) x BM468550 (1-1093)

QY 116 MetProAspAsnLeuArgMetTrpLeuGlnIleTyRAspGlnGluIleGlyLeuLeu 135
||||| : : : : : ||| ||||| |||
DB 903 TTGCCACAGATTCGATTTGGCTCAAG----- 874

QY	136	SerGlyLyAlaGlnLeuSnsGlnGluLeuArgThrGlnGlyValProAlaGluVal	155
Db	873	-----GCAGATTTTATACCCACCGGGTGT-----	847
QY	156	HisAlaLeuMetAspAsnGlyHisAlaHisAsnSprProMetArgTrpAsnGlnGlyTrp	175
Db	846	---GCCAGCTTTTCGAAGGGAGGTTT-----TAAGCTTAT	814
QY	176	-----ProTrpAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAlaTrpThr	192
Db	813	CTTTCCCTCCCTGGCTGGTGGCTTGGCCGATTTGGTCGATAGGCGCTCCAGCANCAACAC	754
QY	193	GlyCysValAlaThr-----AlaAlaAlaGlnIleMetArgTrpHisSer	207
Db	753	AGTTAGTAGCAGACGCCCCCAAGATTGAAAGCTCCCTCCATCCAAAGATCCGAGA	694
QY	208	TrpProLeuGlnGlyLeuGlySerPheAspTrpHisAla-----GlySerIle	223
Db	693	AAACTGATTTAGAGGGGGCCAGAGTACAGAGAACAGAGGGCTGCTGCATGGCCATCG	634
QY	223	uValGlyAsnTrpSer-----GlyThrPheGlyGlyMetTrpAspTrp	237
Db	633	CGTTGGCCCAATGGCGCTCTGTAGAAATACTTGCATATCATGTGTGAG-----AGCTG	580
QY	237	PileAsnMetProGlyAsnProAsnLeuAsnAsnLeuThrGlnSerGlnValAspAla--	256
Db	579	GATTCAAGTTCGAAGGCCAAGCCCAAGACGCTCCAT-----GACTCCATATACAGCGATA	526
QY	257	-----TyrAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTrp	274
Db	525	GTTGCTGTACGCTGTACAGAAAGAGCTGCAATCAACGACCAAGCCGACCA-----	474
QY	274	uAsnGlySerGlyThrTrpSerValTrpValValGlyAlaLeuArgAsnAsnPheArg	294
Db	473	-----GCACCAACCAATCATCATGATCA	451
QY	294	LysArgSerLeuGlnLeuHisValArgAlaLeuTrpThrSerGlnGlnTrpHisAspMe	314
Db	450	TACGACTTCATATTTATTTATTCACCAAGATGCTGTATGAGACGCTCCACCATATCATG	393
QY	314	IleArgGlyGluLeuAlaSerGlyArgProValTrpTyrAlaGlyAsnAsnGlnSerIle	334
Db	392	-----AGCCAAACATTTATGAGAGGATATCCATGACACTCGCTGGTGGCATGGAATA	340
QY	334	egly---HisAlaPheValCysAspGly-----TyrAlaSerAs	346
Db	339	TACCTTCATCTCTTGAAGAAGACAGTAAATTGATTTGGAAATGATAAAGAACCA	280
QY	346	pGlyThrPheHisPheAsnTrpGly-----TrpGlyGlyValase	359
Db	279	TGGAAATGAAGCTTCATATTCACCTCCAGCCCAAGCTCCATCTGGGGGGGTATTC	220
QY	359	TAsnGlyPheTrpLysLeuThrLeuLeuSerProThrSerLeuGlyIle-----	375
Db	219	CAACTGGA-----CCTCCAACTGCTGCTGGTGGCATTTAAAGTAG	181
QY	376	-----GlyGlyGlyGlyIleGlyPheThrIleTrpGlnGluIle-----	388
Db	180	ATAAATTCCAAAATGACAGTCA--ATCCAAATATCGTTATATCGCGGATCCGACAGC	124
QY	389	-----IleThrGlyIleGluProAlaLysThrProAlaGluAlaGlyThrAsp	405
Db	123	CGAGCGTGTCCGAGGAGCTG--GTGGCGCAGCTGAGATTCCACGGGGCCACAGGCTCCGCT	65
QY	405	IaleuProIleLeuAla	410
Db	64	CCCGCTCTTCGCTGCTGCT	48

AK004957 3170 bp mRNA linear HTC 19-JAN-2002

LOCUS Mus musculus adult male liver cDNA, RIKEN full-length enriched

library, clone:1300010G06:glycosylphosphatidylinositol specific phospholipase D1, full insert sequence.

AK004957.1 GI:12836534

HTC; CAP trapper

Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA, clone:1300010G06.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

2 99279253
10349636

3 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, H. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

4 11042159

5 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, T., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawa, J., Okazaki, Y., Muramatsu, H., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

6 20530913
11076861

7 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Frieschmann, M., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakado, I., Pesole, G., Quackenbush, J., Schriml, L.M., Struhl, F., Suzuki, K., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, P., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, R., de Bonaldo, M.F., Brownstein, M.J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustinch, S., Hill, D., Hofmann, M., Hume, D.A., Kamuya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Rinswald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

8 21085660
11217851

9 5 (bases 1 to 3170)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bull, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Haneagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirokawa, T., Horii, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Konda, M., Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, K., Sasaki, N., Sasaki, D.,


```

OY 443 GYTYRARGLeuAnLysAlaAspGluValIleGluValLysThrSerSerIleAsn 462
Db 1746 -----GAC 1748
OY 463 ILeSerTrpTyrGlyTyrGlyLuhHisProGluSerPheSerLeuAlaProAsnGlnLeu 482
Db 1749 TTCTCGTGTGGTAACTCCCTTAC----- 1775
OY 483 SerGlnGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGlnGlnTrpGlu 502
Db 1776 -----GGGTCACAGTGGCAACACAGACCTCTCTGATTTGGAGCCGACCTGGAA 1829
OY 503 ProVal---ArgHisAlaGlnGlyGlyTyrValAsnSerIleLysValAsnThrThrAsp 521
Db 1830 AATGTCAGCAGATGGACCGCTCAACCCACAAAAG----- 1865
OY 522 ProAsnAsnValValValThrValAspAsnAsnGlnGlyLysLeuSerIleValProAsn 541
Db 1866 -----AACCAAGAGGAAAAGACCTGGGAAAGGTGTACGGC 1901
OY 542 SerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPheAsnSer 561
Db 1902 TACTTCTCCCGAAC-----CCGCAAGCACAATCATCATC----- 1937
OY 562 AspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGly----- 578
Db 1938 ---TCGGAGATTAAGCGCATGGGAAAGCTGGTACCTCCCTGCAAGTGTATGTAGAGA 1994
OY 579 -----AlaThrAlaAspAspValIle 585
Db 1995 GTGAAATGGAGCCCTAACCAAGCTCGTGGTGGAGCCCTCATGATGAGCGTGTCT 2054
OY 586 SerLeuGlyTyrValMetAlaGluValProGlyLysSerSerAsnTyrProValValTyr 605
Db 2055 AAAATGGCATTCCTA----- 2069
OY 606 SerLysAspValLeuThrLeuSerGluGlyAspTyrThrLeuTyrThrArgPheSerIle 625
Db 2070 -----ACCATGACCCCTGCCACAGGCGAGGCGACCTCGATGACCAATGACT--- 2117
OY 626 AsnAsnGlnLysAspGluTyrLysLysIleGlySerValSerValLysThrProThrGlu 645
Db 2118 -----CCCGAG 2123
OY 646 TyrThrHisProLeuPheGluValGlnHisAsnGlnThrSerThrTyrThrLeuAsnMet 665
Db 2124 AACACACAGCCTCTCTG-----CTGACGACATTCAGTGGAGACCGA 2165
OY 666 AlaHisAsnArg-----ValLeuProAspPheThrLeuLysAsnGlyLeuPro 682
Db 2166 CGCTTTCTCGATTTGGTAGTGTCTGCACTGACCTGATGATGATGCTTA--- 2222
OY 683 PheAsnGlyLeuLeuValValVal-----PheArgGlnThrGlnSerSerSerGlySer 700
Db 2223 -----GATGAATCATCATGCTGCTGCCACCTGACAGATTAACAGATGTCACCTTGGAGCTG 2276
OY 701 LeuThrAlaAlaGlnGluThrValHisIleLysGlnGly 713
Db 2277 CTAGGGGAGAAAGATGGCGAGTGTACATTTAATAGGC 2315

```

```

RESULT 6
LOCUS AU140457 820 bp mRNA linear EST 05-AUG-2002
DEFINITION AU140457 PLACE3 Homo sapiens cDNA clone PLACE3000100 5', mRNA
sequence.
ACCESSION AU140457
VERSION AU140457.1 GI:11001978
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 820)

```

```

AUTHORS Ota,T., Suzuki,Y., Saito,K., Ishii,S., Yamamoto,J., Sugiyama,T.,
TITLE Nishikawa,T., Nakamura,Y., Sugano,S., Masuho,Y. and Isogai,T.
HRI human cDNA project (Ota,T., Suzuki,Y., Saito,K., Ishii,S.,
Yamamoto,J., Sugiyama,T., Nishikawa,T., Nakamura,Y., Sugano,S.,
Masuho,Y., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end y pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1. 820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE3000100"
/clone_lib="PLACE3"
/issue_type="Placenta"
/notes="Vector: pME18SFL3"
BASE COUNT 218 a 220 c 183 g 195 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 1 Length: 820
Score: 102.50 Matches: 58
Percent Similarity: 37.41% Conservative: 46
Best Local Similarity: 20.86% Mismatches: 89
Query Match: 2.31% Indels: 85
DB: 9 Gaps: 13
US-10-030-330-1 (1-843) x AU140457 (1-820)
OY 412 LysAspAlaGlnAlaGluTyrLysSerGluSerGlyLeuAsnValGly----- 427
Db 82 AGAGGTGCAATTTCTCCCTGACAGACATGACAGAGGCTTAAACAATCTCTGCTGATCA 141
OY 428 -----TyrSerIleTyrAsn-----ThrGly 434
Db 142 GATATGTAGTAGAGTCTCTCCAGTGTCTACGAACAACATGAGACACACCTTGTAGAGA 201
OY 435 GlnGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGluValIle 454
Db 202 AGACAGAAACAGGCTTATATCCCACTGCGCATGACTTTTCTGAT----- 249
OY 455 GluValLysThrSerSerIleAsnIleSerTyr----- 465
Db 250 ---ATTACGCCAACCTTTTACTGTGCACTGGATTTGCTCGAGCCACATCACTGCGC 306
OY 466 TyrGlyTyrGlyLuhHisProGluSerPheSerLeuAlaPro-----AsnGlnLeuSer 483
Db 307 TACAGATGGCCATCATCCGACGACCTCACTGGGAGACCTCGAGAAAGATGGGGGCC 366
OY 484 GlnGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGlnGlnTrpGlu 503
Db 367 CACTCTCGGAATTCACCTCACTCCACCAACCTCACTCCAGAGAG----- 414
OY 504 ValArgHisAlaGlnGlnGlyTyrValAsnSerIle---LysValAsnThrThrAspPro 522
Db 415 -----TATGTGGTCACAGTCGTTGCTCTTAATGGCAGAGAGAA 453
OY 523 AsnAsnValValVal-----ThrValAspAsnAsnGlnGlyLysLeuSerIle 538
Db 454 AGTCCCTATATGATGGCCACACATCAACAGATTTCGATGTCGAGAGGACCTGGAAATT 513
OY 539 ValProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGln 558
Db 514 GTTGCTGCAGCCGCCACGACCTGATGATGAGTGTCTGCTGTCTACAGTAGAGA 573

```

QY 559 PheAsnSerAspSerProAspGluLeuArgThrProValAlaPheAlaLeuSerGlnGly 578
Db 574 TAT-----TACGAGTACTTTACGGA 594
QY 579 AlaThrAlaAspAspValIleSerLeuGlyTrpValMetAlaGluValProGlyGlySer 598
Db 595 GAACA-----GGAGGAAT 609
QY 599 Ser-----AsnTrpProValValTrpSerLysAspValLeuThrLeuSer----- 613
Db 610 AGCCCTGTCAGAGTTCACCTGCTGGAGCAAGCTACACCTACACCTACAGCCCTT 669
QY 614 -----GluGlyAspTrpThrLeuTrpTrpArgPheSerIleAsnGlnLysAspGlu 631
Db 670 AAACCTGGAGTGTATATACATC-----ACTGATATGCTGTCTACCTGCGCTGGAGAC 726
QY 632 TrpLysValIleGlySerValSerValLysThrProThrGluTrpThrHisPro 649
Db 727 CCCGACGACGACGACCAATTTCCATTATACCGAAGCAAGAAATGGCAACCA 780
QY 727 BH770545/c 1594 bp DNA linear GSS 01-MAY-2002
DEFINITION LMGtag308 MG1363 Random Sequence Tag Library Lactococcus lactis
ACCESSION BH770545
VERSION BH770545.1 GI:20373502
KEYWORDS GSS.
SOURCE Lactococcus lactis subsp. cremoris.
ORGANISM Lactococcus lactis subsp. cremoris
Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
Lactococcus.
1 (bases 1 to 1594)
BOLotin,A., Ehrlich,S.D. and Sorokin,A.
Studies of genomes of dairy bacteria Lactococcus lactis
Sci. Aliments, (2002) In press
Contact: Sorokin A
Genetique Microbiome
INRA
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
Tel: 33 1 34 65 25 16
Fax: 33 1 34 65 25 21
Email: sorokine@jouy.inra.fr
best homologue in strain IL403 is ynfH (27%)
Class: shotgun
High quality sequence start: 30
High quality sequence stop: 1566.
Location/Qualifiers
1. 1594
/organism="Lactococcus lactis subsp. cremoris"
/strain="MG1363"
/db_xref="taxon:1359"
/clone_lib="MG1363 Random Sequence Tag Library"
/note="Vector: pSGM2; Site.1: SmaI; Library of
chromosomal fragments of L.lactis strain MG1363 was
prepared by partial AluI digestion or by sonication."
BASE COUNT 362 a 354 c 397 g 466 t 15 others
ORIGIN

Alignment Scores:
Pred. No.: 3.53 Length: 1594
Score: 102.50 Matches: 124
Percent Similarity: 31.70% Conservative: 64
Best Local Similarity: 20.91% Mismatches: 196
Query Match: 2.31% Indels: 209
DB: 17 Gaps: 28
US-10-030-330-1 (1-843) x BH770545 (1-1594)
QY 271 SerPheTrpGluAsnGlySerGlyThrTrpSerValTrpValGluAlaLeuArgAsn 290
Db 1535 TCTCTTTGGCGCGATGNAAGAAAGAGCNCNATATATGCGTGTCTTTTGGCCCTCCGAAT 1476

QY 291 AsnPheArgTrpLysArgSerLeuGlnLeuHisValArgAlaLeuTrpThrSerGlnGlu 310
Db 1475 NATTTCTCCAGAGCTCCNCATGTGA-----AGTCTCTAGTAATAGCCCAT 1425
QY 311 TrpHisAspMet-----IleArg----- 316
Db 1424 TGGTATCCGGTTCAGGAGACTCCCGTCCATTTNCCCNNTCCATATAGATATATAC 1365
QY 317 GlyGluLeuAlaSerGlyArgProValTrpTrpAlaGluAsnGln----- 332
Db 1364 GGAATCCAGTGTGAGACAGGAGCCACAGATGAAAGTGAATTAACAGTACGTCTCCCA 1305
QY 333 -----SerIleGlyHisAlaPheValLysAspGlyTrpTrpAlaSerAspGlyTrpPheHisPhe 351
Db 1304 GGCTCAGTGTGCTTAATGCTCCATTCCTCCATTCACACCAACCAACATAGTTGA----- 1254
QY 352 AsnTrpGlyTrpGlyGluValSerAsnGlyPheTrpLysLeuThrLeuSerProThr 371
Db 1253 -----GATATGTCAGTGAACCCACACCTACAGGACGACAGATCCAGT 1206
QY 372 SerLeuGlyTrpGlyGluGlyIleGlyPheThrIleTrpGlnGluIleIleThrGly 391
Db 1205 CTAGTAGCT-----CCAAACAGT 1188
QY 392 IleGluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeu----- 409
Db 1187 AACTTACCTGCGAACATCTGACGATGGAGCCGACCATCCATGCTCTTCACGCG 1128
QY 410 AlaLeuLysAspIleGluAlaGluTrpLysSerGlySerGlyLeuAsnValGlyTrpSer 429
Db 1127 AGCTCAAGAC----- 1116
QY 430 IleTrpAsnThrGlyGluGlnSerAsnLeuAspLeuGlyTrpArgLeuAsnLysAla 449
Db 1115 -----AGTACAGGA-----GCT 1104
QY 450 AspGlyGluValIleGluValLysThrSerSerIleAsnIleSerTrpGlyTrpGly 469
Db 1103 GACATTCACATCCATCTGATGATTCACATTC----- 1071
QY 470 GluHisProGluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIle 489
Db 1070 -----TACGACACGACGATACCAATGTAATACGATACGATACGATACGATC 1020
QY 490 ThrLeuLeuTrpArgArgThrGlyThrGluGlnTrpGluProValArgHisAlaGlnGly 509
Db 1019 ACT-----TATACGCGCGCAGGA-----TATGACAGCTGACCCAAACG----- 981
QY 510 GlyTrpValAsnSerIleLysValAsnThrThrAspProAsnValValValThrVal 529
Db 980 -----CTGATGTCTCTGACCCAGACGATACAGTCCCCAGAT-----GCTCCAAACGTT 930
QY 530 AspAsnAsnGlnGlyLysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSer 549
Db 929 GGAATGTCCACAGGAATAGAAC-----AATGCTTATACGTTACAGGACGATGCC 879
QY 550 TrpGluHisSerThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThr 569
Db 878 GAACCTAATTCACACATCAATTAAGATTAATATGAGAT----- 837
QY 570 ProValAlaPheAlaLeuSerTrpGlyAlaThrAlaAspAspValIleSerLeuGlyTrp 589
Db 836 -----ACGATGGAAACCGAAGCAGACAGCA----- 810
QY 590 ValMetAlaGluValProGlyGlySerSerAsnTrpProValValTrpSerLysAspVal 609
Db 809 -----AGTGAATATACACAGTACTCTCCAGGCTCAGTT 774
QY 610 -----LeuThrLeuSerGlyLysAspTrpThrLeuTrpTrpArgPheSer 624
Db 773 GGTCTCAGCTCCATTCACATCAATCACTGCAACAGAT----- 738

[illegible]

Db 324 TGGAACGCCATCATCTCTACAGACCAATGCGTTTCCCAATGAAGAACTTCAGCA 383
 Oy 485 -----GlyTleasnThrIleThrleuLeuTyrArgArgThGlyThrIleGlnTrpGluPr 503
 Db 384 GGTGCATTTAAATACGCGAGCGCTTTGTGT-----GACTGCCAGCTGAGATGGCTCC 437
 Oy 503 ovalaTgHlAaInglnIglYrValAsnSerIleValAsnThrThrAspProAs 523
 Db 438 ACAGTGGGTGGCGGAGAACACTTTCAGACTTCGTAAATGCCAGTTGTGCCACCTCA 497
 Oy 523 nasnValValIThrValAspAsnInglnIglYrLysSerIleValProAsnSerPh 543
 Db 498 GGTGTG-----AAAGGAGAGAGATTTTCAGCTGACGCCACAGATGGCTT 542
 Oy 543 eValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPhAsnSerAsp 563
 Db 543 TGTGTGTGAT-----GATTTCACCAACCCAGATACCGCTCCAGCAGAAAGCAGTC 596
 Oy 563 rProAspGluIleArGThrProValAlaPhaAlaLeuSerThrGlnAlaThrAlaAsp 583
 Db 597 GGCAATTAAG-----GGTCCGACGAGATTTTACTGCTGCGCTGACAGCACTGACTC 653
 Oy 583 pValIleSerLeuGlyTTPValMet-----AlaGluValProG 596
 Db 654 CCTATAGCGCTGCTGCGTGAAGAAAGAACAGAACGCCCTTCAGAGTCAGAGATG----- 708
 Oy 596 yGlySerSerAsnTyr-----Pr 602
 Db 709 -----GAGACTACGCGACACCTCGCGGCCAGGAGGCGAGCTGATGATATACAC 761
 Oy 602 ovalaITrpSerIleAspValLeuThrIleSerGlnGly----- 615
 Db 762 CATCTGAGGCTGGCAATGTGCAATACACGAGAGGAGAACAGTGTGTCATCTC 821
 Oy 616 -----AspTyrThrIleuTyrTyrArgPhaSerIleAsnInglnLysAs 630
 Db 822 CAACCACTTGGTCTGCTGATACCTCCGCAAGCCACATCACCATCACTGTCCTTC 881
 Oy 630 pGluTPrLysLysIleGlySerValSerValLysThrProGlnLysTyrThr----- 647
 Db 882 C-----TTTACCAGAACCCCGCATGAGCACTCCACCATCCGTC 917
 Oy 648 -----HisProLeuHeGluValGlyLys 655
 Db 918 CGGGGCCATGCGCCGCTGAGTGTGCTGCGCGGGCACCCAGCCACAGATAGCTG 977
 Oy 655 sAnInglnThrSerThrTyrThrIleuAspMetAlaHisAsnArg-----ValLeuPr 672
 Db 978 GCAGAAAGGATGAGGACGACACTTCCCGCGCCGCGAGAGAGCAGCATGCACTGATGCC 1037
 Oy 672 o-----AspPheThrLeuLysAsnLeuGlyLeuProPheAs 684
 Db 1038 CGAGGATGACGCTTCTTTCATTCGTCGAGAGATGAGAGACATCCGCG----- 1086
 Oy 684 ngIyLysLeuValValAlaPheArgGlnThrGlnSerSerSerGlySerLeuTrpAla-- 703
 Db 1087 -----GTTTACAGCTGCAACAGCTCAGAACAGTGCAGGAGCGCTCTGCAAA 1133
 Oy 704 -AlaInglnThrVal-----HisI 710
 Db 1134 CGCAACCTGACTGCTGATAGACAGCCGCTATTTTTCGCGCCCTGTTAGACAGAACCGT 1193
 Oy 710 eLysInglnGlyThrPheValTyrLysProValGlnGlyProIlePro----- 727
 Db 1194 CACCAAGGAGAAACACCGGCTCTGACAGTCATTCGAGGAGGCCCGCCCTCGGCT 1253
 Oy 728 -----AspGlySerTyrArgAlaThr-----LeuHisAlaPheValAsnG 741
 Db 1254 GAATGACGAGAAAGATACACAGCCCTTTGGTGGTGACCGAGAGGCACTTTTTCAGACAGG 1313
 Oy 741 yGlnInglnIleuTyrLeu-----LysGlyLysArgAsnTyrTh 754
 Db 1314 CAACCACTGCTGTATTCGTCGAGACTGAGAGCTGAGGAGCGCGGAAA-----TACAC 1367

Oy 754 rValLysIleValAsnGlyThrAlaValAlaIleGlySerSerGlnGluIleArgVa 774
 Db 1368 GTGTGAGATGTCACATACCTTGGCACCGAG-----AGAGGACAGCTTGCGCTGACTGT 1421
 Oy 774 lPheProAsnProAlaArGAspTyrValGlnIleSerAlaProCysIle 790
 Db 1422 GATCCCTACCTCCGACCTGCGACTCCCTCACATACATACGCGCCATCGCTG 1470
 RESULT 11
 AK018132
 LOCUS
 DEFINITION
 Mus musculus adult male medulla oblongata cDNA, RIKEN full-length
 enriched library, clone:6330406P08.homolog to KIAA1001 PROTEIN,
 full insert sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE
 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitahara, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, K., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multichannel sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
 Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
 Kuehl, P., Lewis, S., Matsuno, Y., Nika, I., Pesole, G.,
 Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
 Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
 Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bulic, C.,
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
 Hofmann, M., Hume, D. A., Kamliya, M., Lee, N. H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P.,
 Rling, B., Ringwald, C., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
 Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmink, L.,
 Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.
 and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)

TITLE
 JOURNAL

Db 614 GCAGAGTTCGAA-----TTGATAGCCTGACAAATGACA 649
 Oy 173 GINGLYTYRProTPrasAnSLySGLProLeuProLeuAnSLyASnHISAlATyThr 192
 Db 650 GCTCCATAC-----CCAGATACGATCTCGATGCGG 682
 Oy 193 GTCYSValAlaThrAlaAlaInleMeTArgTyRHS-----SerTrpPro 209
 Db 683 GGAGGCTTCTTACTCCAGAGCCAGCAGCGGTTAGATAGGCTCAGTCGCCG 742
 Oy 210 LeuGInGLyGLyGLySerPheAspTyRHisAlaGLySerLeuValGLyAsnTrpSerGly 229
 Db 743 -----CGGCTAGGATTTCTCTCGAAGAGATGATGCG 775
 Oy 230 ThrPheGLyGLyMetTyRAsp---Triple 238
 Db 776 AAGTTGGGTATGCTATGAGGTATGGGTC 805

CHUT 16
 5828
 INSERTION 765 bp mRNA linear EST 05-MAR-2002
 BM815828 EST593922 BNR Medicago truncatula/Meloidogyne incognita mixed EST
 library cDNA clone pBNIR-28F2, mRNA sequence.
 BM815828
 BM815828.1 GI:19151842
 EST.
 Medicago truncatula/Meloidogyne incognita mixed EST library.
 Medicago truncatula/Meloidogyne incognita mixed EST library.
 Eukaryota: mixed EST libraries.
 1 (bases 1 to 765)
 Cheung, F., Tsai, J. and Fraser, C.M.
 ESTs from roots of Medicago truncatula after infection with the
 nematode Meloidogyne incognita
 Unpublished (2002)
 Contact: Bird, DM
 Plant Nematode Genetics Group
 North Carolina State University
 Box 7616, Raleigh, NC 27695, USA
 Tel: 919 515 6813
 Fax: 919 515 9500
 Email: david_bird@ncsu.edu
 TIGR sequence name: MTOBJ25YK More information is available at:
 www.medicago.org
 Seq primer: SKmod (CTA GAA CTA gta gAT CC).
 Location/Qualifiers
 1. 765
 /organism="Medicago truncatula/Meloidogyne incognita mixed
 EST library"
 /cultivar="A17"
 /db_xref="taxon:188702"
 /clone="PBNIR-28F2"
 /clone_1lb="BNIR"
 /tissue_type="3 week old roots"
 /dev_stage="3 days after infection with Meloidogyne
 incognita second stage larvae"
 /lab_host="XtOLR"
 /note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the UniTap XR vector from
 StrataGene and packaged using GigaPack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in XtOLR cells."

BASE COUNT 217 a 174 c 145 g 229 t

Alignment Scores:

Pred. NO.: 2.59 Length: 765
 Score: 98.50 Matches: 54
 Percent Similarity: 31.87% Conservative: 26
 Best Local Similarity: 21.51% Mismatches: 86
 Query Match: 2.22% Indels: 85

DB: 14 Gaps: 9
 US-10-030-330-1 (1-843) x BM815828 (1-765)
 Oy 395 AlAlaYstThrProAlaGLyInleThrAspAlaLeuProIleLeuAlaLeuAspIle 414
 Db 193 GCAGCATTGATTCACGATTGGCTTGAACCCAGCCCTAAATTTGAGACTTCCGCTCA 252
 Oy 415 GluAlaGLyTyRysSerGLySerGLyLeuAsnValGLy----- 427
 Db 253 ATTGAAGCAAGATATATCTATGGTGTGTAAGTTGGATTGATACAACTGCTTCA 312
 Oy 428 TYRSerIleTYRAsnThrGLyGLyGLyInleSerAsnLeuAspLeuGLyTYRGLyAsn 447
 Db 313 TCTCTCAACATTAATTAAGCCGGGATTCCTTCATTAACCAAGATTTCTGCAACACTTATG 372
 Oy 448 LysAlaAspGLyGLyValIleGLyValIleSerSerIleAsnIleSerTrpTyRGLy 467
 Db 373 CTGGCTGAT-----AAGGACAGTCTCTGAAGGATCTTACATTCAT 414
 Oy 468 TYRGLyGLyHisProGLySerPheSerIleuAla-----ProAsnGLySerGLn 484
 Db 415 TATGTGACCCGCCAGATGATTAACAGTCTCTGTAACCTGCTCACAAGTTGCTCT 474
 Oy 485 GLYIleAsnThrIleThrLeuLeuTYRArgTyRGLyThrGLyInleTrpGLyProVal 504
 Db 475 TCTGAGAACATTTATACCTT-----GGAGCTCAGC----- 507
 Oy 505 ArgHisAlaGLyGLyTYRValAsnSerIleValAsnThrThrAspProAsnAsn 524
 Db 508 -----CAATGATGCCAATAAACA 525
 Oy 525 ValValAlaThrValAspAsnAsnGLyGLyLeuSerIleVal----- 539
 Db 526 GTTTGAAGACTCGATTCAGATGATGAGGAAAGCTCCTTCAATGCCAAGCGCTGG 585
 Oy 540 ---ProAsnSerPheValAlaAspLeuAsnSerTYRGLyHisSerThrIleThrValGln 558
 Db 586 AGGCCAAATTCATC-----ATAACCTATCT 612
 Oy 559 PheAsnSerAspSerProAspGLyIleArgThrProValAlaPheAlaLeuSerThrGLy 578
 Db 613 GCCGATGATGATGCCAATAAATCATTTGTCACCTGCCCAATTTGCTGCTCTT--- 669
 Oy 579 AlThrAlaAspAspValIleSerLeuGLYTPValMetAlaGLyValProGLyGLySer 598
 Db 670 -----TCT 672
 Oy 599 SerAsnTyRProValAlaTrpSerIleAsp-----Val 609
 Db 673 CTCAGCCTTAACCTGGTTTGGCAATCAATCTTATTCACAGACTTTTAACTTACCATTT 732
 Oy 610 LeuThrLeuSerGLyGLyAspTyRThrLeuTrp 620
 Db 733 CTATCATTTTACATTGTGTCATTAACCTCTG 765

RESULT 17
 BF383380/c 1378 bp mRNA linear EST 27-NOV-2000
 LOCUS 602045042P1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4194395 5',
 DEFINITION mRNA sequence.
 BF383380
 ACCESSION BF383380.1 GI:11364685
 VERSION
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1378)
 NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: c9agabcs-remail.nih.gov
Tissue Procurement: jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: L1MAG5528 row: d column: 12.

FEATURES	Location/Qualifiers
source	1. .1378

315 a	356 c	362 g	342 t	3 others
ORIGIN				

Alignment Scores:	
Pred. No.:	7.9
Score:	98.50
Percent Similarity:	31.61%
Best local Similarity:	19.96%
Query Match:	2.22%
DB:	12
	Gaps:
US-10-030-330-1 (1-843)	1378
x BF383380 (1-1378)	89
	Conservative:
	52
	Mismatches:
	129
	Indels:
	176
	Gaps:
	19


```

/clone= "BB0AAV0C12"
/clone_1lb="BB0AA"
/note="end . T3"

```


Mon Jun 2 11:44:00 2003

us-10-030-330-1.p2n.rst

Page 22

Search completed: June 2, 2003, 01:02:24
Job time : 2993 secs

Porphyromonas gingivae
Porphyromonas gingivae
Porphyromonas gingivae
Porphyromonas gingivae
Porphyromonas gingivae
Porphyromonas gingivae
Escherichia coli p
Novel human diazotro
Porphyromonas gingivae
Porphyromonas gingivae
Porphyromonas gingivae
Amino acid sequenc
F. balustellum CPD7
Porphyromonas gingivae
Porphyromonas gingivae
Listeria monocytogenes
Mycobacterium gallii
Enterohaemorrhagica
Enterohaemorrhagica
Listeria monocytogenes
Staphylococcus epidermidis
Kx2 endoprotease.
Kx2 protease. Sa
Muramidase release
Novel human diazotro
Bifidobacterium bu
PMGA 1:2 protein c
H. pylori HPN165 p
Protease. Pyrococ
Pyrococcus furiosus
W09855926 Seq ID 6
Amino acid sequenc
Lactococcus lactis
Novel human diazotro
Listeria monocytogenes
Sulfated fucose con
Escherichia coli p

XX Aglus CT, Barr IG, Hocking DM, Margolis MB, Patterson MA;
 PI. Ross BC, Rothel LJ, Webb EA;
 XX WPI: 1999-385613/32.
 DR N-PSDB: AAX91703.
 XX
 PT Antigenic Porphyromonas gingivalis peptides for preventing
 XX gingivitis
 PS Claim 1; Page 466-468; 588pp; English.
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAX94318 to
 CC AAX94583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The Pg polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 SU Sequence 843 AA:
 Query Match 100.0%; Score 4438; DB 20; Length 843;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKSFLAIYMLGIMOGSHAPVTERALSLRLALROVLSMGOTAVSDKSIDYVR 60
 DB 1 MKSFLAIYMLGIMOGSHAPVTERALSLRLALROVLSMGOTAVSDKSIDYVR 60
 QY 61 OGDAGRGITSOEGSPAYFYVANGNNEGATVAADRIPTILAYSPIGRFDMSMPDNL 120
 DB 61 OGDAGRGITSOEGSPAYFYVANGNNEGATVAADRIPTILAYSPIGRFDMSMPDNL 120
 QY 121 RMWLOIYDOEIGILSGKAOLNEELRTGCVPAEVALMDNGHFANDPMKNOGYPMNNK 180
 DB 121 RMWLOIYDOEIGILSGKAOLNEELRTGCVPAEVALMDNGHFANDPMKNOGYPMNNK 180
 QY 181 EPLLPNGNHAHYTCVATAAQRYSWPLQEGSGFDYHAGSLVGNMGSFGEMYPWIMN 240
 DB 181 EPLLPNGNHAHYTCVATAAQRYSWPLQEGSGFDYHAGSLVGNMGSFGEMYPWIMN 240
 QY 241 PCNPDDNLTOGQVADYATLMRDVSAVSMSFYENGSGTYVYVGLARNRFFKRSIQL 300
 DB 241 PCNPDDNLTOGQVADYATLMRDVSAVSMSFYENGSGTYVYVGLARNRFFKRSIQL 300
 QY 301 HVALTTSQEMHDMIRGELASGRVYVYAGNOSIGHAFVCDGASDGTFFHNGKGVSN 360
 DB 301 HVALTTSQEMHDMIRGELASGRVYVYAGNOSIGHAFVCDGASDGTFFHNGKGVSN 360
 QY 361 GFYKLTLSPTSLGIGEGIGFTIYOEITIGIEPARTPAAGYDALPILALKDIEYKS 420
 DB 361 GFYKLTLSPTSLGIGEGIGFTIYOEITIGIEPARTPAAGYDALPILALKDIEYKS 420
 QY 421 ESGLANVGYSTYNTGEOQNLDCYRLNKADEVIETKSSINISWYGYGHPSPSLAPN 480
 DB 421 ESGLANVGYSTYNTGEOQNLDCYRLNKADEVIETKSSINISWYGYGHPSPSLAPN 480
 QY 481 QLSOGINTITLLYRRGTGEOMEVYRHAOGGYVNSIVNTTDPNNVVYVDNNEGKSIYP 540
 DB 481 QLSOGINTITLLYRRGTGEOMEVYRHAOGGYVNSIVNTTDPNNVVYVDNNEGKSIYP 540
 QY 541 NSFVADLNSYEHSTIVOFNSDSPDEIRTPVAALSTGATADYVLSLGMVMAEVPGGSSN 600
 DB 541 NSFVADLNSYEHSTIVOFNSDSPDEIRTPVAALSTGATADYVLSLGMVMAEVPGGSSN 600
 QY 601 YPVVMSKDVLTLSGEGDYTLMYRFSINNOKDEMKKISVSKPTEYTHPLFEVGHNOTST 660
 DB 601 YPVVMSKDVLTLSGEGDYTLMYRFSINNOKDEMKKISVSKPTEYTHPLFEVGHNOTST 660
 QY 661 YTLDMANRVLDPDTLKNLGLPENGELVVVFFROTOSSSGLMAAQTVAHKOGETVYKRP 720

DB 661 YTLDMANRVLDPDTLKNLGLPENGELVVVFFROTOSSSGLMAAQTVAHKOGETVYKRP 720
 QY 721 VVEGPIPDGSRATLAFVNGOOLYLKGRKNTVIVNGTAVALESESEIYVFPNPAR 780
 DB 721 VVEGPIPDGSRATLAFVNGOOLYLKGRKNTVIVNGTAVALESESEIYVFPNPAR 780
 QY 781 DVEISAPCIPORTSITLFDLSGRIMKNSLSAGHRMDVSRPLNCAVILKYDGYTTKIN 840
 DB 781 DVEISAPCIPORTSITLFDLSGRIMKNSLSAGHRMDVSRPLNCAVILKYDGYTTKIN 840
 QY 841 IVH 843
 DB 841 IVH 843
 RESULT 2
 AAX94360
 ID AAX94360 standard; Protein; 849 AA.
 AC AAX94360;
 XX
 DT 25-AUG-1999 (first entry)
 DE Porphyromonas gingivalis protein PG28.
 XX Porphyromonas gingivalis; Pg; periodontal disease; gingivitis;
 KM vaccine; antigenic.
 XX Porphyromonas gingivalis.
 OS MO929870-A1.
 PN
 XX
 PD 17-JUN-1999.
 XX
 PF 10-DEC-1998; 98MO-AU01023.
 XX
 PR 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002311.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003354.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 XX
 PA (CSLC-) CSL LTD.
 XX
 PI Aglus CT, Barr IG, Hocking DM, Margolis MB, Patterson MA;
 PI. Ross BC, Rothel LJ, Webb EA;
 XX WPI: 1999-385613/32.
 DR N-PSDB: AAX91578.
 XX
 PT Antigenic Porphyromonas gingivalis peptides for preventing
 XX gingivitis
 PS Claim 1; Page 323-324; 588pp; English.
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAX94318 to
 CC AAX94583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The Pg polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 SU Sequence 849 AA:

Query Match 100.0%; Score 4438; DB 20; Length 849;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSFLAIYMLFGIMOGHSAPYTERALSLARLALROYSLRMGQTAVSDKISIDYVR 60
 DB 7 MKSFLAIYMLFGIMOGHSAPYTERALSLARLALROYSLRMGQTAVSDKISIDYVR 66
 QY 61 OGDAGERTSOEGSPAYFYVANGNNEGALVAADRITLALSPYIGRFDMSDNL 120
 DB 67 OGDAGERTSOEGSPAYFYVANGNNEGALVAADRITLALSPYIGRFDMSDNL 126
 QY 121 RMWLOIYDOEIGLISGKAOLNBEILRTGEPVPAEVAHLMONGFANDPMRNOGYPNKK 180
 DB 127 RMWLOIYDOEIGLISGKAOLNBEILRTGEPVPAEVAHLMONGFANDPMRNOGYPNKK 186
 QY 181 EPLLPGNHAHTGCVATAAAQIMRYHSMPLQGGSPDYHAGSLVGKNSGTFGEYDWINM 240
 DB 187 EPLLPGNHAHTGCVATAAAQIMRYHSMPLQGGSPDYHAGSLVGKNSGTFGEYDWINM 246
 QY 241 PGMPDLDNLTQSOVDVATLMDRVASVSMSFEYENGSGTYSYVVGALRNFRKRSLOL 300
 DB 247 PGMPDLDNLTQSOVDVATLMDRVASVSMSFEYENGSGTYSYVVGALRNFRKRSLOL 306
 QY 301 HVALYTSQEMHDMINGELASGRPYVYAGNNOISGHAFCVGDYASDGTFFHNMGGVSN 360
 DB 307 HVALYTSQEMHDMINGELASGRPYVYAGNNOISGHAFCVGDYASDGTFFHNMGGVSN 366
 QY 361 GFYKTLTSPSTIGIGEGIGFTIYOEITIGIEPANTPAEGADALPILAKDIEAEKYS 420
 DB 367 GFYKTLTSPSTIGIGEGIGFTIYOEITIGIEPANTPAEGADALPILAKDIEAEKYS 426
 QY 421 ESSLANGSYIYNNGEESNLDGRLNKADGEVIEYKTSINISWYGGHPESFSLAPN 480
 DB 427 ESSLANGSYIYNNGEESNLDGRLNKADGEVIEYKTSINISWYGGHPESFSLAPN 486
 QY 481 QLSQGITNTLLAYRRTGTQEWEVPRHQAQGVNSIKVNTDPNNVVYTVNNEGKLSIVP 540
 DB 487 QLSQGITNTLLAYRRTGTQEWEVPRHQAQGVNSIKVNTDPNNVVYTVNNEGKLSIVP 546
 QY 541 NSFVADLNSYESTIVQENSUSPDEIRFPVAFALSTGATADYISLGWMAEVPGGSSN 600
 DB 547 NSFVADLNSYESTIVQENSUSPDEIRFPVAFALSTGATADYISLGWMAEVPGGSSN 606
 QY 601 YPVVMSKDVLTLSGGYTLMTYRPSINNOKDEMKKIGSVKRTETHPLEFVGHNTST 660
 DB 607 YPVVMSKDVLTLSGGYTLMTYRPSINNOKDEMKKIGSVKRTETHPLEFVGHNTST 666
 QY 661 YTLDMANRVLDPFTLKNLGLPENGELVYVFROTQSSSGSLMAAOETVHIKQGETFYKP 720
 DB 667 YTLDMANRVLDPFTLKNLGLPENGELVYVFROTQSSSGSLMAAOETVHIKQGETFYKP 726
 QY 721 VVEGPIPDGSYRATLHAFVNGOOLYLKGRNTYVKIYNGTAVAEIESSEIRVFPNPAR 780
 DB 727 VVEGPIPDGSYRATLHAFVNGOOLYLKGRNTYVKIYNGTAVAEIESSEIRVFPNPAR 786
 QY 781 DVEISAPCIPOETSIITLFDLSGIYKMSLSAGHGMVSRPNAGYILAKVGYTTKIN 840
 DB 787 DVEISAPCIPOETSIITLFDLSGIYKMSLSAGHGMVSRPNAGYILAKVGYTTKIN 846
 QY 841 IVH 843
 DB 847 IVH 849

RESULT 3
 AAB36100 standard; Protein; 844 AA.
 AC AAB36100;
 AC AAB36100;
 DT 16-FEB-2001 (first entry)
 XX

DE Porphyromonas gingivalis periodontain.
 XX Porphyromonas gingivalis; periodontain; antiinflammatory; antibacterial;
 KW amidolytic; alpha₁-proteinase inhibitor; periodontitis; gingivitis.
 XX Porphyromonas gingivalis.
 XX W0200063394-A2.
 PD 26-OCT-2000.
 PF 20-APR-2000; 2000MO-US10574.
 PR 21-APR-1999; 99US-0130436.
 XX (UYGE-) UNIV GEORGIA RES FOUND INC.
 PA (TRAV/) TRAVIS J.
 PA (POTEMPA) POTEMPA J.
 PA (NELS/) NELSON D.
 PI Travis J, Potempa J, Nelson D;
 XX WPI: 2000-679600/66.
 XX Novel oral bacterial periodontain polypeptide for treating periodontal
 PT diseases, has amidolytic activity for cleavage of non-denatured human
 PT alpha₁-proteinase inhibitor at reactive site loop region of inhibitor
 PT
 XX
 PS
 XX Example 1; Fig 1; 55pp; English.

CC The present sequence is given in a specification relating to novel
 CC oral bacterial polypeptide referred to as periodontain. The polypeptide
 CC has amidolytic activity for cleavage of denatured polypeptides and
 CC non-denatured serpin polypeptides. It has amidolytic activity for
 CC cleavage of a non-denatured human alpha₁-proteinase inhibitor at a
 CC reactive site loop region of the inhibitor. Periodontain is useful for
 CC inhibiting the peptidase activity and reducing periodontitis, loss of
 CC tooth attachment and periodontal pocket formation, and for reducing
 CC growth of bacteria, preferably P. gingivalis in vitro or in vivo.
 CC It is useful for protecting an animal from a disease caused by
 CC P. gingivalis and for treating periodontal diseases, including
 CC gingivitis and periodontitis.
 CC
 XX

Sequence 844 AA:
 Query Match 88.9%; Score 3946.5; DB 21; Length 844;
 Best Local Similarity 89.8%; Pred. No. 4,66-289;
 Matches 758; Conservative 4; Mismatches 81; Indels 1; Gaps 1;

QY 1 MKSFLAIYMLFGIMOGHSAPYTERALSLARLALROYSLRMGQTAVSDKISIDYVR 60
 DB 1 MKSFLAIYMLFGIMOGHSAPYTERALSLARLALROYSLRMGQTAVSDKISIDYVR 60
 QY 61 OGDAGERTSOEGSPAYFYVANGNNEGALVAADRITLALSPYIGRFDMSDNL 120
 DB 61 OGDAGERTSOEGSPAYFYVANGNNEGALVAADRITLALSPYIGRFDMSDNL 120
 QY 121 RMWLOIYDOEIGLISGKAOLNBEILRTGEPVPAEVAHLMONGFANDPMRNOGYPNKK 180
 DB 121 RMWLOIYDOEIGLISGKAOLNBEILRTGEPVPAEVAHLMONGFANDPMRNOGYPNKK 180
 QY 181 EPLLPGNHAHTGCVATAAAQIMRYHSMPLQGGSPDYHAGSLVGKNSGTFGEYDWINM 239
 DB 181 EPLLPGNHAHTGCVATAAAQIMRYHSMPLQGGSPDYHAGSLVGKNSGTFGEYDWINM 240
 QY 240 PGMPDLDNLTQSOVDVATLMDRVASVSMSFEYENGSGTYSYVVGALRNFRKRSLOL 299
 DB 241 PGMPDLDNLTQSOVDVATLMDRVASVSMSFEYENGSGTYSYVVGALRNFRKRSLOL 300
 QY 300 LHVVALYTSQEMHDMINGELASGRPYVYAGNNOISGHAFCVGDYASDGTFFHNMGGVSN 359
 DB 301 LHVVALYTSQEMHDMINGELASGRPYVYAGNNOISGHAFCVGDYASDGTFFHNMGGVSN 360

Oy	360	NGEYKLTLLSPSLGIGGEGIEFTTYOEITIGIEPAKPPAEAGTALPLTALKDIEAEK	411
Dd	361	NGFYKLTLLSPSLGIGGEGIEFTTYOEITIGIEPAKPPAEAGTALPLTALKDIEAEK	420
Oy	420	SESGLNAGSYIYNTGEEOSNLDLGYRLNKADAEVIEVKTSINISWYGYGHEPESPFLAP	479
Dd	421	SESGLNAGSYIYNTGEEOSNMDLXYRLNKADAEVIEVKTSXXNISWYGYGHEPESPFIAP	480
Oy	480	NOLSGINTITLLYRTGTGEOMEPRRHAGGIVNSIKVNTTPDNNVYTVDDNNEGLSTY	539
Dd	481	NOLSGINTITLLYRTGTGEOMEPRRHAGGIVNSIKVNTTPDNNVYTVDDNNEGLSTY	540
Oy	540	PNSFPAADINSYEHSTITVOFNDSDEIERTPAFALSTGATADVDVSLGMYAAEYVGSS	599
Dd	541	PNSFPAADINSYEHSTIXYQFNDSDEIERTPAFALSTGATADVDVSLGMYAAEYVGSS	600
Oy	600	NYPVVWSKDVLTLSGDDYTLWYRFSINNOCKDEMKKIGSVSVKTPREYTHPLFEVGHQTS	659
Dd	601	NYPVVWSKDVLTLSGDDYTLWYRFSINNOCKDEMKKIGSVSVKTPREYTHPLFEVGHQTS	660
Oy	660	TYTLDMAHNRVLPDFTLKNLGLPENGELVYVVROROSSGSLMAOETVHIQGEFFYVK	719
Dd	661	TYTLDMAHNRVLPDFTLKNLGLPENGEXYVVRQTOSSXGSLMAOETVHXKOGEFFYVK	720
Oy	720	PVVEGPIDGSRATLHAFFVNGQOOLYLKGKKNYTVKLYNGTAVEAIESSSEIRVPNPA	779
Dd	721	PVVEGPIDGSRATLHAFFVNGQOOLYLKGKKNYTVKXXNNGTAVEAIESSSEIRVPNPA	780
Oy	780	RDYVEISAPCIPOETSIILFDLSGRIYMKNSLSAGHGRMDVSRPLPGAVILKVDGYTKI	839
Dd	781	RDYVEKSAPICIPOEXXIIIFDXSGRIYMKNSLSAGHGRXDVSRLPGAVIIXKVDGYTKI	840
Oy	840	NIVH 843	
Dd	841	NIVH 844	
RESULT 4			
AAB36099			
ID	AAB36099	standard; Protein: 840 AA.	
XX	AAB36099;		
AC			
XX			
DT	16-FEB-2001 (first entry)		
XX			
DE	Porphyromonas gingivalis proteinase prtr.		
	Porphyromonas gingivalis; prtr; periodontal; antiinflammatory;		
	antibacterial; amidolytic; alpha ₁ -proteinase inhibitor; periodontitis;		
	gingivitis.		
XX			
OS	Porphyromonas gingivalis.		
XX			
PN	WO200063394-A2.		
XX			
PD	26-OCT-2000.		
XX			
PE	20-APR-2000; 2000MO-US10574;		
XX			
PR	21-APR-1999; 99US-0130436.		
XX			
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.		
PA	(TRAV/) TRAVIS J.		
PA	(POTE/) POTEPA J.		
PA	(NELS/) NELSON D.		
XX			
PI	Travis J, Potempa J, Nelson D;		
XX			
DR	WPI: 2000-679600/66.		
PT	Novel oral bacterial periodontal polypeptide for treating periodontal		
PT	diseases, has amidolytic activity for cleavage of non-denatured human		

PI	alpha1-proteinase inhibitor at reactive site loop region of inhibitor	
PT	-	
PS	Example 1; Fig 1; 55pp; English.	
XX		
CC	The present sequence is given in a specification relating to novel	
CC	oral bacterial polypeptide referred to as periodontan. The polypeptide	
CC	has amidolytic activity for cleavage of denatured polypeptides and	
CC	non-denatured serpin polypeptides. It has amidolytic activity for	
CC	cleavage of a non-denatured human alpha1-proteinase inhibitor at a	
CC	reactive site loop region of the inhibitor. Periodontan is useful for	
CC	inhibiting the peptidase activity and reducing periodontitis, loss of	
CC	tooth attachment and periodontal pocket formation, and for reducing	
CC	growth of bacteria, preferably <i>P. gingivalis</i> in vitro or in vivo.	
CC	It is useful for protecting an animal from a disease caused by	
CC	<i>P. gingivalis</i> and for treating periodontal diseases, including	
CC	gingivitis and periodontitis.	
XX		
SQ	Sequence 840 AA:	
	Query Match 24.3%; Score 1078; DB 21; Length 840;	
	Best Local Similarity 34.4%; Pred. No.2.e-72;	
	Matches 301; Conservative 97; Mismatches 402; Indels 74; Gaps 25;	
QY	1 MKKSF-----LAIWLFGLAMOGHSAPTKRALSLAR--LALROYSLKQGTAVSDK 52	
DB	1 MKRIFFYTLGLDILCLPML-----QAGPVTRSAEQAKNFKAPKROPTLLSSXAS----- 49	
QY	53 ISIDVYRQGAERKGTISQEEGSPAYFVVARGNNEGALVAAADRIPTILAYSPRIGFD 112	
DB	50 XRXDXVYAAEREE-----ALFVFNRGEXKGFVLVLAADDRFEXXGYAFKHFED 99	
QY	113 MDSMPDLRMWLQIYDDEIGILISGKAOLNEILR--TEGVPAEVHAIADNGHPANDPM 169	
DB	100 AARPDNLRGWLKCKYKREMLXMGKAPRIPRIEAKPTRPXPSISIAFXXTGCHASDPI 159	
QY	170 RMNGVYPMNNKEPLLPGNNHAYTCGVATAAQAQIMRHSYMPLOGSGSPDYHAGSLVG---N 226	
DB	160 LMXQGYPTNLTLPPLPSSQQAITYGCVATAMGQIMHNTWPKASGXHDY--DDMTGHTH 218	
QY	227 WSGTFGEYWDIMNPNPDLNLTLQSOYDAYATIRDYSASVSMSEFNGSGTYSVYVVG 286	
DB	219 XSGTFGEYWNMSKMPGNISY--GXSPXYKAKASTFKADVSVFVNQGFXXFGSTPSXHYER 277	
QY	287 ALRNNFRKRSQQLHVALYITSOEWHDIRBELASGRVYTYAGNNOSIGAFVCDGYASD 346	
DB	278 ALRETFHFKSLRIYHRSILRGKEMXMDIRRELLENPPVYVYAGDASGNAHFVCDGYKPD 337	
QY	347 GTFHFMQMGVSNCFYKLTLLSPSTLAI--GEGSIGFTIYOEITIGIEPARTPAEAGDA 405	
DB	338 GTFHFMQMGVSNCFYKLTLLSPSTLAI--GEGSIGFTIYOEITIGIEPARTPAEAGDA 405	
QY	406 LPILAKDIEAEYKSEGLNAYSIYNTGEEOSNLDGLRYLNKADGIEVKTSSINISW 465	
DB	396 DPTXLYIGXQ--HNMSDEALDXSVIKINSTYAGDKXKILYRLTLPLNGTETPAPVAXXPYIW 454	
QY	466 YG-YGEHNEFSIAPNOLSOGINITLLYKRTGTGEQWPRVNAHAGGYVNSIKVWTPDPNN 524	
DB	455 EDIIGESTGNIXIPISOFAEGKNTLXLYRFDGAMWMEKXNIIILGLVKNKLEV--TMPAGD 513	
QY	525 VVWYVDNNEGKLSIYPNFVADLNSVEHSTITVQFNSDPSPEIRTPVAFALSTGATADV 584	
DB	514 VAIYS--ADGKLYAKKDSGLSHDKAKYSDCKKXAIAYVINGTKEPFSRYTFALRN--TEBRX 569	
QY	585 ISLGWMAEV--PGGSSNTPVYVWSKDVLTLSGDTYLTWRSINNOKDE--WKITGSVSK 641	
DB	570 YFLGRHXHVEIHPGDEGEKYSILXITGLKARAGQYMLVCTGMEIMEDASVIEKXSKEVA 629	
QY	642 TPTEYTHLFEVGNHNSFTFLDMAHNNVLDPTLKNL--GLRFGELVYVVRQOQSSGS 700	
DB	630 EHTSTHSSLVANSNQDILLXKRNANPPTLPTFSKYNDEGGATFGSKKEXVAIKAFSET--F 688	
QY	701 LMAQETHIKQGETFVYKPVVEGP-----IPDGSYATLHAIVNGQOOLY----- 746	

Db 689 FQAKREHXSQACETKVLSPEXTANSXYTNALFPDGRY-----YIYAREGGFMDPXD 742
QY 747 LAGKRYTYKRYNGTAVEAIESE--EIRYFPNADYVEISAPCIPOETSTILFDLSCK 804
Db 743 LFGDYRYRXXTTDLSSDIAGNDVSTIYXXPNPADHYVXKAIPTYAGXXKXFLFDIOGR 802
QY 805 IYKKNLSAGHGMDVSRLLPNCAYILKVDGYTTK 838
Db 803 MOXSTXKESADMDRVDYERLPKGIYIYVEDMVGK 836

RESULT 5

AA81812
ID AAY81812 standard; protein; 398 AA.

AC AAY81812;

XX 09-JUN-2000 (first entry)

S. pyogenes cysteine protease speB/ protein sequence.

XX Cysteine protease; speB; Group A Streptococcus; extracellular protease;
XX detection; diagnosis; extracellular matrix; infection; skin infection;
XX disease status monitoring; vaccine; Streptococcus mediated disease;
XX pharyngitis; tonsillitis; scarlet fever; sepsis; erysipelas; fasciitis;
XX pneumonia; acute rheumatic fever; poststreptococcal glomerulonephritis;
XX cellulitis; bacteraemia; meningitis.

OS Streptococcus pyogenes.

XX US6030835-A.

XX 29-FEB-2000.

XX 16-SEP-1997; 97US-0931220.

XX 02-DEC-1993; 93US-0160965.

XX 14-SEP-1994; 94US-0306542.

XX (BAYLOR COLLEGE MEDICINE.

XX Kapur V, Musser JM;

XX WPI: 2000-205208/18.

XX N-PSDB; AAA07111.

XX Determining the presence of a Group A Streptococcus expressing a
protease capable of degrading proteins of the extracellular matrix,
using a specific antibody.

PS Disclosure: Column 7-8; 56pp; English.

XX This sequence represents the S. pyogenes cysteine protease speB7.
XX The invention relates to a method for determining the presence of a
XX Group A Streptococcus which expresses an extracellular protease
XX (preferably speB) capable of degrading proteins of the extracellular
XX matrix, comprising: (1) combining a sample with an assay medium
XX comprising a first member of a specific binding pair which binds to a
XX second member of the binding pair to form a complex, where the first
XX member has at least 1 epitopic site competitive with at least 1 conserved
XX epitopic site on the protease; and (2) detecting complex formation as
XX indicative of the presence of the pathogenic organism. The method is
XX useful for screening host samples for evidence of infection with
XX Streptococcus pyogenes and for monitoring the disease status of the host.
XX The speB gene products are used as a vaccine for protecting against
XX Streptococcus mediated diseases such as pharyngitis, tonsillitis,
XX skin infections, scarlet fever, sepsis, erysipelas, fasciitis, pneumonia,
XX acute rheumatic fever, poststreptococcal glomerulonephritis, cellulitis,
XX bacteraemia, and meningitis.

XX Sequence 398 AA;

Query Match 11.1%; Score 494; DB 21; Length 398;
Best Local Similarity 30.6%; Pred. No. 8,9e-29;
Matches 129; Conservative 73; Mismatches 155; Indels 64; Gaps 13;

QY 6 LIAIVMLFGIAMOG-----HSAPVTKERALSLARLALROYSLRMGTAVSDKISIDYV 59
Db 10 LLSLALGFEVLNPNYAFADQNFARNEREKADSAITFIQKSAATKAGARSAD-1KLDKVN 68
QY 60 RQDAERGITSQEBGSPATFYVANRGNNEGIALVADDRIPITLAIYSPIGRFMDSPDN 119
Db 69 LGSEL-----SGSNMYVYNISTG--GFVYVSGDKRSPILLGYSTGSFDAQ-KEN 116
QY 120 LRMWLOIY-----DOEIGLILSGKAOINEEILRTGCVPAVHALMDNGHFPANDPRMNQ 173
Db 117 IASFMEYVQIENKRLDTTYAGTAIRQPV-----VSLSD-----SKGIHQY 162
QY 174 GYPWNKEPLLPN-----GNHAYTGVATPAAOIMRYHSNPLQEGSFYD----- 218
Db 163 GNPYNLTLPVIEKVKPGEQSFVGQHAATGCVATATQAQIMRYHNPKNGLKDYTTLSNN 222
QY 219 ----HAGSLVGNNSGTFGEKMYDMIN-MGPNPDLNLTQSDVAVATLMRDVSAVSMSFY 273
Db 223 PYFNHPKRL--FPAISTROYNNMNTLPYSGRESNVQKM--ATSELMADVGISVDMDYG 277
QY 274 ENGSGTYVYVYVAGALNNFRYKRSLOLHVRLATYSQEHMDIRGELASGRPYVYAGNQS 333
Db 278 PSSGSGASSRNVQALAKNFQYINOSVHQINSRDSKQDEMOAIDKELSONQPYVYQGVKV 337
QY 334 IGAHAYVCDGYASDGTFFHNNMGVSNGFYKLLSLFSLGIGEGIGFTTYOETITGIE 393
Db 338 GGHAFVIDGADGRNFYHVNMGMGVSDGPFRLALNPSALGTGCGAGFNGYSAVYGIK 397
QY 394 P 394
Db 398 P 398

RESULT 6

ABP29579 standard; protein; 398 AA.

XX ABP29579;

XX 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 8334.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus pyogenes.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

XX Tettelein H;

XX WPI: 2002-352536/38.

XX N-PSDB; ABN70210.

PT New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 PS Claim 1; Page 3948; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (II), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 XX Streptococcus proteins.

XX Sequence 398 AA;

Query Match 11.1%; Score 493; DB 23; Length 398;
 Best Local Similarity 30.6%; Pred. No. 1.1e-28;
 Matches 129; Conservative 72; Mismatches 156; Indels 64; Gaps 13;

OY 6 LLAIVMLFGIAMOG-----HSAFVTKERALSIALRLAROVSLRMQAVSDKISIDYY 59
 DB 10 LLSLALGFLVLANPVFADQNFARNEKEAKDSATFPIORSAIKAGARSAED-IKLDKYN 68
 OY 60 ROGDAERGITSQEGSPAYFYANRGNNGYALVADDDRLPTLAVSPGRFDMSPDN 119
 DB 69 LGGEL-----SGSNMTVYINISG--GFVYVSDKSPETLGYSTGSGSPANG-KEN 116
 OY 120 LRMLQIY-----DOELGILISGKAOLNEELIRTEGVPAEVALDNGHFANDPWRMNQ 173
 DB 117 IASFMSYVEQJKEKKLDITYAGTAEIKQPV-----VKSLLD-----SKGIHYNQ 162
 OY 174 GYPMNKEPLRN-----GNHAYTCVATAAQAIRYHSMPLQSGSPDY----- 218
 DB 163 GNPYNLLTFVIEKVRKQSGFVGQHAATGCVATATQIKYHNPYPMKGLDYTYTLSSNN 222
 OY 219 ----HAGSLVGNMSTGFEEMWDIN-MPGPDLDNLTQSOVADYATLMDVSAVSMSFY 273
 DB 223 PTFNHKKNL---FAAISTROYNNNTLPTYSGRHSNVOKM--AISELMADVGISVMDYG 277
 OY 274 ENGSGTYVYVYGALRNFRYKRSLOLHVALYTSQEMHDMIRGELASGRPVYAGNNOS 333
 DB 278 PSSGASGSSRYOALKENFGYNOSVHQIRKGFDSKQDWEAQIDKELSQNPVYVGQVKV 337
 OY 334 IGHAFVCDGYASDCTGFHFPMWGVNGYKLTLLSPYSLGIGEGICGTIYOELLITGIE 393
 DB 338 GGHAFVTDGADGRNFYHVMWGVSDGFFRLDALNPSALGTCGGAGGNGYOSAVVGIK 397
 OY 394 P 394
 DB 398 P 398

RESULT 7
 AA07898
 ID AA07898 standard; Protein; 398 AA.

AC AA07898;
 XX
 DT 22-JUL-1997 (first entry)
 XX
 DE Streptococcus pyogenes clone speB7 pre-pro cysteine protease.
 XX Pre-pro; cysteine; protease; inhibition; neoplastic; proliferation;

KW cell; human; treatment; carcinoma; sarcoma; melanoma; lymphoma;
 KW leukemia; leukemia; blood; lung; mammary gland; prostate;
 KW intestine; stomach; liver; heart; skin; pancreas; brain tissue;
 KW wound covering; prevention; metastasis; identification; speB7.

XX Streptococcus pyogenes.

XX Key Location/Qualifiers

FT Misc-difference 216

FT Domain /note="corresponding codon TAG"
 333..338
 /label= nucleotide_binding_domain

XX MO9634941-A1.

XX 07-NOV-1996.

XX 30-APR-1996; 96WO-US05997.

XX 01-MAY-1995; 95US-0432692.

XX (BAYU) BAYLOR COLLEGE MEDICINE.
 (TEXA) UNIV TEXAS SYSTEM.

XX Ananthaswamy HN, Fernandez A, Kapur V, Musser JM;
 PI WPI: 1996-506148/50.
 XX N-PSDB: AAT45219.

XX Use of extracellular Streptococcal cysteine protease enzyme - for
 PT inhibiting the proliferation of neoplastic cells, e.g. for treating
 PT carcinoma, lymphoma or leukemia.

XX Disclosure: Pages 59-61; 99pp; English.

CC The present sequence is the Streptococcus pyogenes clone speB7
 CC pre-pro cysteine protease (CP), which can be used to inhibit
 CC neoplastic cell proliferation, especially in a human, useful in the
 CC treatment of neoplastic conditions, e.g. carcinomas, sarcomas,
 CC melanomas, lymphomas and leukemias originating from blood, lung,
 CC mammary gland, prostate, intestine, stomach, liver, heart, skin,
 CC pancreas or brain tissue. The CP is especially associated with a
 CC wound covering, and can also be used to prevent metastasis or
 CC identify susceptible neoplastic cells.
 CC K1735 and CM519 melanoma cells were injected s.c. into nu/nu mice,
 CC optionally followed by i.p. injection of CP (100 microl, 24 hours
 CC later). The mice were checked twice weekly for tumour growth for
 CC 12 weeks, to give results that showed that treatment with CP
 CC completely protected athymic mice against transplanted K1735
 CC melanoma growth, and protected 60% of the mice from developing
 CC CM519 melanomas.

SO Sequence 398 AA;

Query Match 11.0%; Score 489; DB 17; Length 398;
 Best Local Similarity 30.4%; Pred. No. 2.1e-28;
 Matches 128; Conservative 73; Mismatches 156; Indels 64; Gaps 13;

OY 6 LLAIVMLFGIAMOG-----HSAFVTKERALSIALRLAROVSLRMQAVSDKISIDYY 59
 DB 10 LLSLALGFLVLANPVFADQNFARNEKEAKDSATFPIORSAIKAGARSAED-IKLDKYN 68
 OY 60 ROGDAERGITSQEGSPAYFYANRGNNGYALVADDDRLPTLAVSPGRFDMSPDN 119
 DB 69 LGGEL-----SGSNMTVYINISG--GFVYVSDKSPETLGYSTGSGSPANG-KEN 116
 OY 120 LRMLQIY-----DOELGILISGKAOLNEELIRTEGVPAEVALDNGHFANDPWRMNQ 173
 DB 117 IASFMSYVEQJKEKKLDITYAGTAEIKQPV-----VKSLLD-----SKGIHYNQ 162
 OY 174 GYPMNKEPLRN-----GNHAYTCVATAAQAIRYHSMPLQSGSPDY----- 218
 DB 163 GNPYNLLTFVIEKVRKQSGFVGQHAATGCVATATQIKYHNPYPMKGLDYTYTLSSNN 222

XX Travis J, Potempa J, Nelson D;
 XX WPI; 2000-679600/66.
 DR Novel oral bacterial periodontal polypeptide for treating periodontal
 XX diseases, has amidolytic activity for cleavage of non-denatured human
 PT alpha1-proteinase inhibitor at reactive site loop region of inhibitor
 PT
 XX Example 1; Fig 1; 55pp; English.
 XX
 XX The present sequence is given in a specification relating to novel
 CC oral bacterial polypeptide referred to as periodontin. The polypeptide
 CC has amidolytic activity for cleavage of denatured polypeptides and
 CC non-denatured serpin polypeptides. It has amidolytic activity for
 CC cleavage of a non-denatured human alpha1-proteinase inhibitor at a
 CC reactive site loop region of the inhibitor. Periodontin is useful for
 CC inhibiting the peptidase activity and reducing periodontitis, loss of
 CC tooth attachment and periodontal pocket formation, and for reducing
 CC growth of bacteria, preferably P. gingivalis in vitro or in vivo.
 CC It is useful for protecting an animal from a disease caused by
 CC P. gingivalis and for treating periodontal diseases, including
 CC gingivitis and periodontitis.
 CC
 CC Sequence 398 AA;
 SQ
 Query Match 10.1%; Score 449; DB 21; Length 398;
 Best Local Similarity 32.5%; Pred. No. 2.2e-25;
 Matches 188; Conservative 53; Mismatches 157; Indels 56; Gaps 13;
 QY 26 KERALSLARLALRQVSLRMGQAVSDKISIDVYRQDAERGITSQEGSPAYFYVANG 85
 DB 36 KEAKDSAXTFIQSAIKAGARSAED-IKLDKYNLGGEL-----SGSNMYVYINISG 86
 QY 86 NNEGVALVADDRIPITLAYSPIGRDMSMPDNLMMLQITDOETGLILSGAQLNEEI 145
 DB 87 ---GFYIVSGDKRSPETLIGYSTGSPDANG-KEXIASFESYVEQI-----KENKRL 134
 QY 146 LRTGEPAP-----VALMDNGHFANDPRMNGVPMN---NKEPLLPN-----GNHAY 191
 DB 135 DTTVAGTAEIKOPXVSLD-----SKGHYNQNGFNILTPRXKXEVKRGDSFVQGH 189
 QY 192 TGCVATAAQIMRYSWPLQEGSGFDY-----HAGSLVGNMSTGTEGMYDIN-M 240
 DB 190 TCCVATATQIMKYNHXPKGLKXYTYTLSSNPNYFNHPKRLXIXIS---TRQYNNMIL 246
 QY 241 PGNPDNDLTQSOVDAYATLMDVSYASVMSPTENGSGTYSYTYVCAALNNRYKSL 300
 DB 247 PTYSGRESNVOKM-AKSELMDVGISVMDXXPPSSGASGSRVQALXENFGYNOSXXQ 304
 QY 301 HVRALTSQEMHDMINGELASGRPVYAGNNSIGHAFCYCDGYASDGTFFHFMNGMGVSN 360
 DB 305 INRSDDXKDDMEQITKELSONPNVYQGVGKYGAFYIDGAXKGNFHXVNMNGMGVSD 364
 QY 361 GFYKTLTSLPTSLGIGESIGFTTYQETITGIEP 394
 DB 365 GFHLDALNPSALGTGCGAGGFGYQSAVXGKIP 398
 RESULT 10
 AA134573
 ID AAY34573 standard; Protein; 540 AA.
 XX
 XX AAY34573;
 XX
 XX 25-AUG-1999 (first entry)
 XX Porphyromonas gingivalis protein Pg91.
 DE Porphyromonas gingivalis; Pg; periodontal disease; gingivitis;
 XX Porphyromonas gingivalis; Pg; periodontal disease; gingivitis;
 KM vaccine; antigenic.
 XX

OS Porphyromonas gingivalis.
 XX
 XX W09929870-A1.
 XX
 XX 17-JUN-1999.
 XX
 XX 10-DEC-1998; 98W0-AU01023.
 XX
 XX 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 XX
 XX (CSLC-) CSL LTD.
 XX
 XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 DR N-PSDB; AAX91791.
 XX
 XX WPI; 1999-385613/32.
 XX
 XX Claim 1; Page 569-570; 588pp; English.
 PS
 XX
 XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91889 represent PCR primers used in the
 CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The Pg polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 CC
 CC Sequence 540 AA;
 SQ
 Query Match 4.3%; Score 190; DB 20; Length 540;
 Best Local Similarity 20.5%; Pred. No. 1.3e-05;
 Matches 103; Conservative 78; Mismatches 162; Indels 160; Gaps 23;
 QY 413 DIEAEKSSGSLNV-----GY-----STYNTGEOSNLDGLRLKADGEVIEVT 458
 DB 101 DIVVTGKNSDIIKIMSVELMNRPGYKSHVAFSRDANQNNKLVYKEDFSNVQIYVDI 160
 QY 459 SSINISWYCGHEPESFSLAPNOLSGITITLLVRRGTEDQMEPRHAGGVNSIKYN 518
 DB 161 AS-----NY-----RSPSLNNGNPPLAFYTF----- 186
 QY 519 TTDPNNVVTVDNNEKSLIVNSFYADNLSYEHSTITVQFNSDSPDEIRTPVAFALSTG 578
 DB 187 -----NNTFKISFY--DYVFLSNGGQNFKNLKFQDDEKKID--KVDLSIG 229
 QY 579 ATADYISLGWMAEV-----PGGSNTPV-----WS-----K 607
 DB 230 STSESMGNHAMPGLMGVPEMNRKQSGSDIGFLSNFYDNDPEROWSGPIKVSFSDMSFSK 289
 QY 608 DVLTSEGDYTL-----WYRFST-----NNQXDEMKIGSVSVKTPREYHPLFEVGHNT 658
 DB 290 IOMLDEDDNNTINGESCHNFMTTYSDDYSEYSDW-----DIRYTPKSKRYEKG 339
 QY 659 STYTLDMANRVLPDF--TLKNIGLPF-----NGELVYVFRQTOSSGSL--MAAEVTVH 709
 DB 340 KTFPTMDLVLEALFLTASYOSETNSGLCYDKNANHLYITTYKKRENGNTLTKYRANVADKH 399
 QY 710 IKQ--GEFF-----YKRPVEGPIPDGSTRATLHAFLVNGQOOL-----YLGKRN 752

DB 400 NNDLMSDFTYTSANALYTPQVD-----INPTKGLVCSWVEYLPGKRI 444
QY 753 Y--TYKIVNGTAVEAIESE-EIRVFPNARDYVEISACIPQETSIIILFDLSGKIVMKN 809
DB 445 VMSDQWTHANGVEDIVMOEGSMKLYPNPAQEAIVISLP-TAANCAKAVYDMQGRVAAE 503
QY 810 SLSAGHRDVSRLPNCAYILKV 832
DB 504 SFGSNEYRLNVOHLAKGTIILKV 526

RESULT 11

AA34430

ID AAY34430 standard; Protein; 563 AA.

AC AAY34430;

PF 25-AUG-1999 (first entry)

Porphyromonas gingivalis protein PG91.

Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
vaccine; antigenic.

Porphyromonas gingivalis.

W09929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98W0-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

PA (CSLC-) CSL LTD.

PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
Ross BC, Rothel LJ, Webb EA;

WPI; 1999-385613/32.

N-PSDB; AAX91648.

Antigenic Porphyromonas gingivalis peptides for preventing
gingivitis

Claim 1; Page 404-405; 588pp; English.

AAX91536 to AAX91801 encode two hundred and sixty six antigenic
Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
activity with a vaccine mechanism of action. The Pg polypeptides can be
used as vaccines especially against Porphyromonas gingivalis. Probes can
be used to detect Porphyromonas gingivalis in standard hybridisation
assays. Porphyromonas gingivalis is involved in periodontal disease
especially gingivitis.

SQ Sequence 563 AA;

Query Match 4.3%; Score 190; DB 20; Length 563;

Best Local Similarity 20.5%; Pred. No. 1.4e-05;

Matches 103; Conservative 78; Mismatches 162; Indels 160; Gaps 23;

QY 413 DIEAEKSESGLVN-----GX-----SIYNTGEQSNLDLGRKLNKADGEVIEVKT 458
DB 124 DIVYTKNMSDRIKIVVELMKNRPGKSKVAAPSRANQAKLYKEDPSNQLDVQI 183
QY 459 SSINISWYIGGEHESFSLAPNQLSGINTITLLYRTGETGEWEPVRAAGGVNSIKVN 518
DB 184 AS-----NY-----RSPSLNNGGNPFALAFAYTGF-----209
QY 519 TDDPNVYVYVONNECKLIVPNSFVADINSVESHITVQFNSDPDELRTVPALFSLNG 578
DB 210 -----NNTKISFV--DYFSLNGGONFNKLNLFSDGKKID--KVDLSLG 252
QY 579 ATADYISLGWVAEY-----PGGSSNYPVY-----WS-----K 607
DB 253 STSESNGHANAPLMGVFEMNKGKSDIGFLSNFVNDPEFOMSGPIKVSSEDSMSFSK 312
QY 608 DVLTLESGDYTL---WYRFSI---NNQDEMKKIGSVSKTPPEYTHPLFEVGHNOT 658
DB 313 IOMLDEDNNTINGESCHNFMITYSDYDEYSDM-----DIRVYPPKSKFYKKG 362
QY 659 STYTLDMANNRVLPDF--TLKNGLPF-----NGLVVFPRQOSSGSL--WAAQETVH 709
DB 363 KPTTMDLVEALFTASQSETNSGLGYDKNANHLYLTITVAKKEENGNTLYKRWANVYDKH 422
QY 710 IKO--GETP-----VYKPVVEGPIPDGSYRATLHAFFVNGQOL-----YLKGRN 752
DB 423 NNDLMSDFTYTSANALYTPQVD-----INPTKGLVCSWVEYLPGKRI 467
QY 753 Y--TYKIVNGTAVEAIESE-EIRVFPNARDYVEISACIPQETSIIILFDLSGKIVMKN 809
DB 468 VMSDQWTHANGVEDIVMOEGSMKLYPNPAQEAIVISLP-TAANCAKAVYDMQGRVAAE 526
QY 810 SLSAGHRDVSRLPNCAYILKV 832
DB 527 SFGSNEYRLNVOHLAKGTIILKV 549

RESULT 12

AA34522

ID AAY34522 standard; Protein; 921 AA.

AC AAY34522;

PF 25-AUG-1999 (first entry)

Porphyromonas gingivalis protein PG57.

Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
vaccine; antigenic.

Porphyromonas gingivalis.

W09929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98W0-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

(CSLC-) CSL LTD.

PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
Ross BC, Rothel LJ, Webb EA;

XX WPI: 1999-385613/32.
 DR N-PSDB: AAX91740.
 XX
 XX Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 PS
 XX Claim 1: Page 511-513; 588pp; English.
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAX94318 to
 CC AAX94583. AAX91802 to AAX91889 represent PCR primers used in the
 CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The Pg polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.

Sequence 921 AA:

Query Match 3.6%; Score 158.5; DB 20; Length 921;

Best Local Similarity 17.7%; Pred. No. 0.0073; Mismatches 242; Indels 327; Gaps 37;

Matches 146; Conservative 111; Mismatches 242; Indels 327; Gaps 37;

XX 221 GSLVGNMSTGFEEMTWIMNP---GNPDIDNTQSOVDAYATLMDVSAVSMSFYENG 277
 DB 215 GONVGRLTWNYPEDY---QPEGKNEEL-QLSGYNIYANGTLAQIK---DVSILEYVD 266
 OY 278 GTYSYVVGALNNFPRKRSLOLH--VRLALYTSQEHMDIRGLASGRVYTAAGNOSTIG 335
 DB 267 STYS-----LRDN-----PLOWECVAVY-----DESTE 291
 OY 336 HAFVCD--GYASDGFHFMMGVSNGEYKLTLSPTSLGIGSEIGFTIYOELITGIE 393
 DB 292 SSTVCGTLATDALIYENFENGSPVNGM-----LVIDADODGFSWGHYL----- 336
 OY 394 PAKTFACTDALPILALDIEAEKSEGLANGY---SIYNTGEGSNLDLGYRLNKA 449
 DB 337 -----NAYDAFP-----GHNGCHSLASAYPGICGPTPNYLTIPKPV 374
 OY 450 DG-EVIEVNTSSINISMYGEGHPESESLAPNOLSGINTITLLYRRTGEO---W--E 502
 DB 375 EGAKRYKYVSTODAMMAA-----EHTAVMASTGTAVDDEVILFEETMTAKPTGAMYER 429
 OY 503 PYRHAOG-GYVNSIKVNTD-----PNNVYVTVNNCKLS-I 538
 DB 430 TINLEPGTKYIAMRHVNCIDYFLKDDITVGTAPSEPEPTDFFVSLIENKGRKLKNY 489
 OY 539 VNSNF-----VADLNSYEHSTTTVOGNSDSDPEIKTPVAFALSTGATADVIS 586
 DB 490 YENGYEPDPTDKDPLQLAGYNIYANGSLVHIODP-----TVLEYIDETYSRDDOYE 543
 OY 587 L-----GWMVAEVPGGSSMYPV 604
 DB 544 VEYCTAVYVNDNIESQVCDKLIYDSQSIILIEGFEAGSIPGELLADAGGNVNM----- 600
 OY 605 MSKDVLTLESGDYTLWYRFSINNOXDEMKIGS-----VSXKTPREY-THPLFEVG--- 654
 DB 601 -----DYYPWTWG-----HDSKCIASPSYLPIMIGVLTDPNIVLTPRLBEAKIV 645
 OY 655 ---HNOSTYTLDMAHNRVLPDFTLKNIGLPPNGELVVVVFOTOSSSSSLNAADETVHI 710
 DB 646 KYVSAQDAVYSAAE-HYAVMSTT---GTAVE-DEVLLFEETMTAKANGAMYERTITL 698
 OY 711 KGEFTF-----YKPVVEGPIPDGSYATL----- 735
 DB 699 PAGTYVIAMRYADCTDMFELLDDITVYKSTETVEPEPYDFFVSLIENKGRKLKNYPNG 758
 OY 736 -----HAFVNGGOOLYK----- 748
 DB 759 YEPDKTDDKKPLQLTGYNITANGSLVHIODPTVLEYIDETYSRDDOYEMEYCTAVYN 818

OY 749 -----GKRNTYKIVNGTAVAEIESSEIRVFPNPARDYVEISAPCIPOETSIILE 799
 DB 819 DNIESQVCDKINLT-----TSLDNIGSDTSLKIPNPASVYVRLEG-LSRSKSTIELY 872
 OY 800 DLSCKIVMKNISLSAGHGRDVSRRLPNGAIIILVDC-----YTKIKIN 841
 DB 873 NALGICILRETHSEKTEIDVSRLDNGVYLIRVGVGNKTTTKEVEI 918

RESULT 13

AAX34521
 ID AAX34521 standard; Protein: 922 AA.

AAX34521;

25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG57.

KW Porphyromonas gingivalis; Pg; periodontal disease; gingivitis;

XX vaccine; antigenic.

OS Porphyromonas gingivalis.

PN WO9929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98AU-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.

PI Agius CT, Barr IG, Hocking DM, Margetis MB, Patterson MA;

PI Ross BC, Rothel LJ, Webb EA;

DR WPI: 1999-385613/32.

DB N-PSDB: AAX91739.

XX Antigenic Porphyromonas gingivalis peptides for preventing

PT gingivitis

PS Claim 1: Page 509-511; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic

CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAX94318 to

CC AAX94583. AAX91802 to AAX91889 represent PCR primers used in the

CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial

CC activity with a vaccine mechanism of action. The Pg polypeptides can be

CC used as vaccines especially against Porphyromonas gingivalis. Probes can

CC be used to detect Porphyromonas gingivalis in standard hybridisation

CC assays. Porphyromonas gingivalis is involved in periodontal disease

CC especially gingivitis.

SO Sequence 922 AA:

Query Match 3.6%; Score 158.5; DB 20; Length 922;

Best Local Similarity 17.7%; Pred. No. 0.0073; Mismatches 242; Indels 327; Gaps 37;

Matches 146; Conservative 111; Mismatches 242; Indels 327; Gaps 37;

OY 221 GSLVGNMSTGFEEMTWIMNP---GNPDIDNTQSOVDAYATLMDVSAVSMSFYENG 277

DB 216 GONVGRLTWNYPEDY---QPEGKNEEL-QLSGYNIYANGTLAQIK---DVSILEYVD 267

QY 605 WSKDVLTLSEGDYTLVRESINNOXDEMKKIGS-----VSKPTPEY-THPLFEVG--- 654
Dy 605 -----DYPPTMTG-----HSEKCIASPSYLPIMIGVLPDNYLTPRLLEGAKLV 649
QY 655 ----HNQSTYTLDMAHNRVLPDFTLKNLGPENGELVYVFPOTOSSSGSLMAAQTVAH 710
Dy 655 ----HNQSTYTLDMAHNRVLPDFTLKNLGPENGELVYVFPOTOSSSGSLMAAQTVAH 710
Dy 650 KYWVSAQDAVYSAE-HYAVMST-----GTAVE-DFVLLFETMTAKANGAMYERTITL 702
QY 711 KQGETF-----VYKPVVEGPIPDGSYRATL----- 735
Dy 703 PACTKIAMRHVCTDMFELLDDITVYRSTETVEPVTDVSVLIENKGRLLKMPNG 762
QY 736 -----HAFVNGQOOLYTK----- 748
Dy 763 YEPDKTDDKPLQLTGYNIVANGSLVHIQDPVLEYIDETYSRRDQVEMECYTAAYN 822
QY 749 -----GKRYTYKIVNGTAVEAIESEERFVFPNPARDVVEISAPICPOETSIILF 799
823 DNIESQSVCDKLNVTI-----TSLDNISQDTSIKITPNPASVYVREG-LSRSKSTIELY 876
800 DLSGKIYMNKSLSAGHRMDVSRPLPNGAVIILKVDG---YTKRINI 841
877 NALGICILRETHSEKTEIDVSRNDGVYLIKRVGKNKTTEKEVEI 922
Db

RESULT 15
AA34392
ID AAY34392 standard; Protein: 938 AA.

XX AAY34392:

DT 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PG57.

XX Porphyromonas gingivalis; PG: periodontal disease; gingivitis;
vaccine; antigenic.

XX Porphyromonas gingivalis.

PN W09929870-A1.

XX 17-JUN-1999.

PF 10-DEC-1998: 98MO-AU01023.

XX 04-AUG-1998: 98AU-0005028.

XX 10-DEC-1997: 97AU-0000839.

XX 31-DEC-1997: 97AU-0001182.

XX 30-JAN-1998: 98AU-0001546.

XX 10-MAR-1998: 98AU-0002264.

XX 09-APR-1998: 98AU-0002911.

XX 23-APR-1998: 98AU-0003128.

XX 05-MAY-1998: 98AU-0003338.

XX 22-MAY-1998: 98AU-0003654.

XX 29-JUL-1998: 98AU-0004917.

XX (CSLC-) CSL LTD.

PI Agius CT, Barr IG, Hocking DM, Margetta MB, Patterson MA;
PI Rose BC, Rochel LJ, Webb EA;

XX MPI: 1999-385613/32.

XX DR N-PSDB: AAX91610.

XX Antigenic Porphyromonas gingivalis peptides for preventing
XX gingivitis

XX Claim 1: Page 361-363; 588pp: English.

CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to

CC AAY34583. AAX91802 to AAX91889 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
SQ Sequence 938 AA:

Query Match 3.6%; Score 158.5; DB 20; Length 938;
Best Local Similarity 17.7%; Pred. No. 0.0075;
Matches 146; Conservative 111; Mismatches 327; Gaps 37;

QY 221 GSLVNMSTGFGEMDWINMP---GNPDIDNTQSGVDVATLMDVSAVSSEFENG 277
Dy 232 GQNVGRLLTNNYEDY-----QPEKGNEL-QLSGNIVANGLLAQK---DVSILEYVD 283
QY 278 GTSYVYVVALRNNFRKSLDLH--VRLYTSQEMDMIRGLASGRVYAGNQSIG 335
Dy 284 STYS-----LRDN-----PLQVEYCVTAAY-----DESIE 308
QY 336 HAFVCD--GYASDGFHFNMGGVNGFYKLTLSPTSLGIGEGIFTYQETITGIE 393
Dy 309 SSTVCGTLHATDAILYENFENGPPVNGM-----LVTDAGDGSMSCHYL----- 353
QY 394 PAKTPAAGTDLPLALKDIEAEKSEGLNNGY---SIYNTGEOSNLDLGYRLNKA 449
Dy 354 -----NAYDAFP-----GHNGHCSLSASVPGIGPVTNPNTLITPRV 391
QY 450 DG-EVLEKTSINISWYIGEHPEFSAPNOLSGITITLLKRRGTQ---W-E 502
Dy 392 EGAKRVKVVSTODAMAA---EHYAVMASTTGAVDFFVLEFETMTAKPAGAWER 446
QY 503 PVRAAG--GYNSIKVNTD-----PNNVYVVDNNEKLS-I 538
Dy 447 TNLPEGTXYIARRHNTCTDIYFLKADITVGTAPSEPEVTDVSVLIENKGRLLKYN 506
QY 539 VPNSF-----VADLSYEHSTITVOFNSDSPDIRTPVAFALSTGATADVIS 586
Dy 507 YPNGYEPDKTDDKPLQLAGYNIVANGSLVHIQDP-----TVLEYIDETYSRRDQV 560
QY 587 L-----GVYMAEVPGGSSNYPV 604
Dy 561 VEXCVTAAYVNDNIESQSVCDKLIYDSQSDIILYEGFAGSIEGWLIDADGDNNVM--- 617
QY 605 WSKDVLTLSEGDYTLVRESINNOXDEMKKIGS-----VSKPTPEY-THPLFEVG--- 654
Dy 618 -----DYPPTMTG-----HSEKCIASPSYLPIMIGVLPDNYLTPRLLEGAKLV 662
QY 655 ----HNQSTYTLDMAHNRVLPDFTLKNLGPENGELVYVFPOTOSSSGSLMAAQTVAH 710
Dy 663 KYWVSAQDAVYSAE-HYAVMST-----GTAVE-DFVLLFETMTAKANGAMYERTITL 715
QY 711 KQGETF-----VYKPVVEGPIPDGSYRATL----- 735
Dy 716 PACTKIAMRHVCTDMFELLDDITVYRSTETVEPVTDVSVLIENKGRLLKMPNG 775
QY 736 -----HAFVNGQOOLYTK----- 748
Dy 776 YEPDKTDDKPLQLTGYNIVANGSLVHIQDPVLEYIDETYSRRDQVEMECYTAAYN 822
QY 749 -----GKRYTYKIVNGTAVEAIESEERFVFPNPARDVVEISAPICPOETSIILF 799
Dy 836 DNIESQSVCDKLNVTI-----TSLDNISQDTSIKITPNPASVYVREG-LSRSKSTIELY 889
QY 800 DLSGKIYMNKSLSAGHRMDVSRPLPNGAVIILKVDG---YTKRINI 841
Dy 890 NALGICILRETHSEKTEIDVSRNDGVYLIKRVGKNKTTEKEVEI 922

RESULT 16
AAG98256

QY	328	AGNNQSGIC-HAFVCDGYASDGTFFHFMG-----WGVSNPFYLLTLSPIS-----	372
Db	438	TGGMVEVKKNNYTTTIGAVAGVGNLISDQKRFVSONITFLGKASGIGTLNLMDATSSPDT	497
QY	373	----LGIGEGEI-----GFTIYQE---IITGTEPAKTEAAGTDPLPILALKDIAEYKS	420
Db	498	VGINVGANGSSGIIVANSKCATLANSIGYGTGGNASKGIGVINISTDSLMLKTSSTAQL--	555
QY	421	ESGLNVGSIYNTGSEEDSNLDGYRLNRADEVEIYEKTSI-----NISWYGEHP	472
Db	556	--LQVG--VLGTGE-----LNTTGGIYKARDTQIALNDKSKGDVAVDQNSLL	600
QY	473	ESFSLAPQLOSGINTITLLYRRT---GTEQM---EPV-----RHAQGYV	512
Db	601	ETFMNMYG--TSGTGTTLLTNNGTLNVEGGEYILGFEPVAGTGLNIGAHGMAADAGFI	658
QY	513	-NSIKV-----NTDPN---NVVYTVDNNEGKLSIYPNSFVADLNSYEHST	554
Db	659	TNAIKVEFGLCGEVFEVFNHTNNSDAGYVDMLITGDDKDKYI-----HDA	704
QY	555	ITVQNSDSPDEIRTPVAFALSTGA--TADVISLGWMAEY---PGSSSNYPVWMSKD	608
Db	705	GHVFEFNGNTYSGKTLVNDGLLTASHRADGYTGMG--SEVTIANPG-----TLD	753
QY	609	VL--TLSGDTYTLWRES---INNCKREKKIGISVSKTPREYHPLFEYGHNOTSYTT	662
Db	754	ILASTNSAGDVTLTNALKGDLIMRYOLSSSKMPEFTATGTEFA---GVAQLDSTFT	809
QY	663	LDMAHNRVLPDFTLKNIGLPENGELVYVFEROTSSSSGSLMAAQEIVHIKOGTFYV---	718
Db	810	LERONTALATHAMLOS-----DSENTTSYKKGEOISIGL-----AMNGCTIIFDDI	856
QY	719	--KEPVEGPIDGSGYRATLHAFFVNGOQQLYLKGRNRYTVKIVNGTAVAEIESSEIRYEP	776
Db	857	PAATLAEGLI-----SVDTLVYGAGDYTWKG--RNYQ---VNGTGDVLID-----VP	898
QY	777	NPARD 781	
Db	899	KPMND 903	

RESULT 17

ABG11747

ID ABG11747 standard; Protein: 2652 AA.

XX ABG11747;

XX AC

XX 18-FEB-2002 (first entry)

XX DE

XX Novel human diagnostic protein #11738.

XX KW

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS

XX Homo sapiens.

XX WQ200175067-A2.

XX PN

XX 11-OCT-2001.

XX PD

XX 30-MAR-2001; 2001WO-US08631.

XX PF

XX 31-MAR-2000; 2000US-0540217.

XX PR

XX 23-AUG-2000; 2000US-0649167.

XX PA

XX (HYSE-) HYSEQ INC.

XX PI

XX Drmanac RT, Liu C, Tang YT;

XX DR

XX WPI: 2001-639362/73.

XX DR

XX N-PSDB: AAS75934.

XX PT

New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 XX biodiversity

Claim 20; SEQ ID No 42106; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences. ABG00010-ABG30377 represent novel human amino acid sequences. The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp://ipd.int/pub/published_pcl_sequences.

Sequence 2652 AA;

Query Match 3.48; Score 153; DB 22; Length 2652;
 Best Local Similarity 17.88; Pred. NO. 0.1;
 Matches 186; Conservative 145; Mismatches 334; Indels 382; Gaps 50;

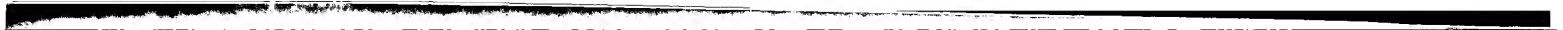
46 QTAVSDKISIDYVYRQC-DAERGITSQEGSPAYFVANRGN-----NEGY 90
 1684 KTAKGKGVCTVLTLPDGTAEADVINEGOTIDIFTAKPGTYIVYFGGVDPINSPF 1743
 91 ALVAADDRITLTAISPGRFMDSPDNLRLML-----QYDEIELI 134
 1744 TVMATEGEV-TAVEEAPV-----MACPGFPMVTEEAIVPSDNGGCGFEPDLVPEA 1797
 135 L-----SGRAQNEIELRTEG-----VPAEV-----HALDNGHFANDPMRW 171
 1798 VRKGEITGEVHMPSCGTATPEIVDKDGTIVRYAFTPEVGHENIKRYGSHIESPLQF 1857
 172 NOGYPMNNK-----EPILPNG--NHAYTCVATAAQAQIMRYHSMPLQEGSFYHAG---- 221
 1858 YVNYNNGSGSVAYGGLVYGVANKTAFTTIEDA-----GEGGLDLAIEGFSK 1906
 222 ---SLVGMMSGTF-----GE-----MYDMINMGNP-----DLDLNLTQOSOVDAV 257
 1907 AEISCIDMKDGTCTVYVLTPLPGDYSLIVKYNOKHIGSPFTAKTIDDSRCSQVRLGSA 1966
 258 ATLNRDVS-----ASVSMSPFENGSGTYSYVVGALRNNEFYKRSLOLHYRALYTQEMND 313
 1967 AFLDLISEIDSLTAST-KABSGRDEPCLRLRPN-----HIGISFIRENGE 2016
 314 -----MIRGELASGRPYVYAGNNSIGHAFPCDGTASDGFHFNW 353
 2017 HLVSIRKKNHNVANSPVIMVQSEIGDARRARYVGRGLSEGRTEMSPFIYDTR--DA 2073
 354 GMGCVS-----NGFYKLTLL-----SPYSLGIG 376
 2074 GYGCSILAEVGSKDIDTEDELDGCKVSTPEYVGVYIVSTRPADHNVGSPFYAKIS 2133
 377 GEGIGTITIOETITGIEPAKTPAEAGTDA-----LPIALDIEAKYESGGLNVGYSIY 431
 2134 GEG-----RVKESITRTSRAPVATVGSICDLAKIPETINSSDMSAHYTSBSG----- 2181
 432 NTGEGOSNLDGRLANKADGEVIEKYSINISWYGEHPEPSLAPNDLSOGINTITL 491
 2182 -----RVTEA--EIVPM-----GKNSHCVRFPDEK--GVHTVSV 2212

492 LVR-----RTGTEQWEP-----VRHAGGYVNSI 515
 2213 KYRGONHTGSPFOETVCPDIEGGAHNRKVRAGPGLEREGAGVPAEFSITWREGAG---GL 2269
 516 KVNITDDNNVYVTVDNNE-GKLSIVPNSFVA-DLNSYEHSITVOPNSDSDPEIRTPVAF 573
 2270 SIAGEVSKAEIIFDDHKNGSCGV---SYIAQEPGVTE--VSIFN---DEHIESPY 2319
 574 ALSTGATADVVISLGMVMAEVRGSSNRYVNSKVDLTLSEGDYTLMTFRFSINNOKDEWK 633
 2320 LVPVIAPSDARLITWASIQESGLKVNOPAS-----FAILNNAK----- 2359
 634 KISVSVKPTETETHPLEFVGNQTSYTLDMANRVL-PDTLKNGLPENGELV--- 689
 2360 --GKIDAKVHS---PSGAVECHVSELEPPKAVRFLPHENGVTIDVFNESHVVGSP 2413
 690 ---VFROQSSGSLMAAQETVHAKOGEFVYRP-----VREGP----- 725
 2414 FVVRVGEFGQAGNPAVSAVGT-GLEGGTGIOSEFFINTTRAGPGLTSTVIEGSPKVKM 2472
 726 ---IPDSYRATLHAFFVNGOOLYLK-GKRNRYVKIYNGTAVAELESSEIRVFPNPAR 780
 2473 DCQETPEG-YKVITYTPMARGNYLISVAKGPNH---TVGKSPKAKVGTQRL----- 2520
 781 DYVEISAPCIPOETSIILFDLSGKIYVK-----NSLSAGHR----- 817
 2521 ---VSPGSAPNETSSILVESVTKVYKKTKYVXAROPKGFIPKPKVTSXGGGSGQKAEV 2576
 818 ---MDVSRLEPNCAYILIKVDGYTT 837
 2577 GQKSSFLVDCSKAGSNMLLIGVHGPTT 2603

RESULT 18
 AAY34479
 ID AAY34479 standard; Protein; 821 AA.
 AC AAY34479;
 XX
 DT 25-AUG-1999 (first entry)
 DE Porphyrymonas gingivalis
 XX Porphyrymonas gingivalis protein PG21.
 KW Porphyrymonas gingivalis; PG; periodontal disease; gingivitis;
 XX vaccine; antigenic.
 OS Porphyrymonas gingivalis.
 XX
 PN WO9929870-A1.
 PD 17-JUN-1999.
 XX
 PF 10-DEC-1998; 98MO-AU01023.
 XX
 PR 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 XX
 PA (CSLC-) CSL LTD.
 XX
 PI Agius CT, Barr IG, Hocking DM, Margetis MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 XX WPI: 1999-385613/32.
 DR N-PSDB; AAX91697.

XX Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 XX
 PS Claim 1; Page 460-461; 588bp; English.
 XX
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The Pg polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 CC
 SQ Sequence 821 AA;
 Query Match 3.3%; Score 148.5; DB 20; Length 821;
 st local similarity 19.3%; Pred. No. 0.035;
 tches 151; Conservative 87; Mismatches 249; Indels 297; Gaps 37;
 176 PMNKEPLIPNGNHATGCVATAAQIMRHS-----W-----208
 216 PLTPDGAYRAGNHTVOGAT-----RYNANNVDLNRNFKDDVAGDHPDGKPMOPEAT 267
 209 ---PLOGEGSF---DYHAGSLVGNMSTGFEYMDIIMPGNDLNLQSOVD-----255
 268 AFMDLEGNTSFVLGAINHGTEYVN-----YPM-----DNKERRHADDDEMYKL 310
 256 ---AVATLMRDVSAVSMSFEYENGSGTYS---YVVVGA LR---NNFRYKRSIQLMHVAL 305
 311 ISRNVAACQASISASMTS---ETNSGILNGSDMYVIRGSDQDANFHRRLRETTLEI---365
 306 YTSQEHMDIRGELASGRPVYTAGNNSGTHAFVCCGYASDGFHFNMGVSGVNGFYKL 365
 366 ---SNMKLVPAQLPRKYNMLNKESTL-----L 368
 366 TLTPSLGIGG-----EGIFTYOETITG---IEPAT-----397
 389 ALIEESLYGHIHGVTAANGOPKLCQILLENHDKRSDVYSDTFTGYVAPITAGYTVK 448
 398 -PAEACTDALPILALDIAEYKSES-GLNVGSIYNTGEEQSNLDLGYRLNKADGEVIE 455
 449 YKAEYPEATRTITDKETVIMDLGNSVPLPVPDFASPMPTISGEVQPOD-----Q 504
 456 VRTSSINISWYCGEPESESLAPNQLSOCIINTITLLRTGTGEQNEPVRAHO-GGYVNS 514
 505 TTNMPTNMEMTEEGQGP-----AMSTEONPLVSYSHRGQYDVT 542
 515 IKV-NTTDPNNVY---VTVDNNEKLSIYP-NSFVADLNSYHSHITTVQFNSDPEIR 568
 543 LKVMANSGSNTITKEFTIVN-----AVMPAAEFGTPEIEGO-TVSFQNSTVAIN 595
 569 TPVAFALSTGATADYISLGWMAEYVGGSSNTPYVMSKDVLSLBSGDYLMFRFSINNQ 628
 596 YWIFEGGTPATSEED-----ENPTVLXSK---AGQYDVL-----627.
 629 KDEMKTIGS--VSVKTPLEYTHLFEVGNHQSITYLDMAHNVLDPFTIKNLGLPENG 666
 628 ---KAISASGETVTKERY-----ITVKRAPVP---A 653
 687 LVVVFQOTSSSSSLAAOETVIHKGEPFVKP-----VEGGRIP-----727
 654 PVADFGTIPRK-----VKKGVTYTFKDLSTNNFTSLWLMFEGGSPATSTEQNP 701
 728 ---DGYRATLAFVNGOOLYLKGRNRYTKIVNGTAVAISSSEIRIRFPNPAR 781
 702 VVTYNTEKRYDQVLTINEGSN--VKKADY-IEVILDSVEDIYAQIGIVIRPQNGTK 758
 782 YVEISAPCIPOETSIILFDLSKTIYKMSLSAGHGRADVSRPLNGAVILKV---DGYTT 837
 759 QILIEANNAIK--AIVLYDINGRVVLKTPNQLRSTVDLSILDEGIYTTINIKTEKSARTE. 816

XX
 OY 838 KINI 841
 DB 817 KIH 820
 817 KIH 820
 RESULT 19
 ID AAY34354 standard; Protein; 869 AA.
 XX AAY34354;
 AC AAY34354;
 XX 25-AUG-1999 (first entry)
 DE Porphyromonas gingivalis protein PG21.
 XX
 XX Porphyromonas gingivalis; Pg; periodontal disease; gingivitis;
 KW vaccine; antigenic.
 XX
 OS Porphyromonas gingivalis.
 XX
 XX W09929870-A1.
 XX
 PD 17-JUN-1999.
 XX
 PF 10-DEC-1998; 98WO-AU01023.
 XX
 PR 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 22-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003358.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 XX
 PA (CSLC-) CSL LTD.
 XX
 PI Agius CT, Barr IG, Hocking DM, Margel's MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 XX
 DR -WPI; 1999-385613/32.
 XX
 DR N-PSDB; AAX91572.
 XX
 PT Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 XX
 PS Claim 1; Page 316-318; 588bp; English.
 XX
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The Pg polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 CC
 SQ Sequence 869 AA;
 Query Match 3.3%; Score 148.5; DB 20; Length 869;
 Best Local Similarity 19.3%; Pred. No. 0.038;
 Matches 151; Conservative 87; Mismatches 249; Indels 297; Gaps 37;
 176 PMNKEPLIPNGNHATGCVATAAQIMRHS-----W-----208
 264 PLTPDGAYRAGNHTVOGAT-----RYNANNVDLNRNFKDDVAGDHPDGKPMOPEAT 315
 209 ---PLOGEGSF---DYHAGSLVGNMSTGFEYMDIIMPGNDLNLQSOVD-----255



•
•
•
•

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:24:52 ; Search time 18 Seconds
(without alignments)
1377.974 Million cell updates/sec

Title: US-10-030-330-1

Perfect score: 4438
Sequence: 1 MKKSFLLATVLMFGIMQGH.....PNCATILKVDGYTTKINIVH 843

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents A.A.
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/prodata/1/1aa/PCITUS.COMB.pep.*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	494	11.1	398	US-08-931-220-5	Sequence 5, Appl1
2	494	11.1	398	PCT-US95-11723-5	Sequence 5, Appl1
3	494	11.1	398	PCT-US96-05997-1	Sequence 1, Appl1
4	140	3.2	934	US-08-840-466A-19	Sequence 19, Appl1
5	140	3.2	934	US-09-696-188B-19	Sequence 19, Appl1
6	140	3.2	10182	US-09-134-001C-3159	Sequence 3159, Ap
7	138	3.1	1398	US-08-750-532-9	Sequence 9, Appl1
8	138	3.1	1398	US-08-894-818B-8	Sequence 8, Appl1
9	138	3.1	1398	US-09-445-472-6	Sequence 6, Appl1
10	128.5	2.9	915	US-09-206-942-35	Sequence 35, Appl1
11	128.5	2.9	1232	US-09-206-942-37	Sequence 37, Appl1
12	128.5	2.9	1238	US-09-206-942-34	Sequence 34, Appl1
13	128	2.9	931	US-08-624-655A-2	Sequence 2, Appl1
14	127.5	2.9	679	US-08-913-942-15	Sequence 15, Appl1
15	127.5	2.9	679	US-09-268-347-26	Sequence 26, Appl1
16	125	2.8	741	US-08-277-231A-4	Sequence 4, Appl1
17	125	2.8	741	US-08-473-750-7	Sequence 7, Appl1
18	125	2.8	741	US-08-477-326-7	Sequence 7, Appl1
19	125	2.8	1168	US-08-620-717A-9	Sequence 9, Appl1
20	124	2.8	1341	US-08-296-791-3	Sequence 3, Appl1
21	124	2.8	1541	PCT-US95-10661A-3	Sequence 3, Appl1
22	123	2.8	621	US-08-328-961-2	Sequence 2, Appl1
23	123	2.8	621	US-08-462-397-2	Sequence 2, Appl1
24	123	2.8	1004	US-09-268-347-30	Sequence 30, Appl1
25	121	2.7	1167	US-08-485-568A-6	Sequence 6, Appl1
26	121	2.7	1167	US-08-590-554A-6	Sequence 6, Appl1
27	121	2.7	1167	US-09-184-223-6	Sequence 6, Appl1

28	120.5	2.7	992	US-09-206-942-61	Sequence 61, Appl1
29	120.5	2.7	998	US-09-206-942-59	Sequence 59, Appl1
30	120.5	2.7	1338	US-08-728-470-9	Sequence 9, Appl1
31	120.5	2.7	1338	US-08-719-641-9	Sequence 9, Appl1
32	120.5	2.7	1566	US-08-687-956A-23	Sequence 23, Appl1
33	120.5	2.7	1599	US-08-617-697-9	Sequence 9, Appl1
34	120	2.7	956	US-09-134-078-63	Sequence 63, Appl1
35	120	2.7	1198	US-09-199-637A-405	Sequence 405, Appl
36	119	2.7	3696	US-09-134-001C-5080	Sequence 5080, Ap
37	118	2.7	2411	US-09-268-347-36	Sequence 36, Appl
38	117.5	2.6	2137	US-09-134-001C-4463	Sequence 4463, Ap
39	117	2.6	1222	US-08-682-517-15	Sequence 15, Appl1
40	117	2.6	1252	US-08-682-517-9	Sequence 9, Appl1
41	116.5	2.6	869	US-08-483-101-15	Sequence 15, Appl
42	116.5	2.6	4302	US-09-052-469-8	Sequence 8, Appl1
43	116.5	2.6	4339	US-09-052-469-6	Sequence 6, Appl1
44	116	2.6	1336	US-08-551-356-6	Sequence 6, Appl1
45	116	2.6	1336	PCT-US93-12687-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1
US-08-931-220-5
Sequence 5, Application US/08931220
Patent No. 6030835
GENERAL INFORMATION:
APPLICANT: Musser M.D., James M.
APPLICANT: Kapur M.D., Vivek
TITLE OF INVENTION: Methods and Compositions for Identifying
TITLE OF INVENTION: Group A Streptococcus
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: WEIL, GOTTSAL & MANGES
STREET: 2882 Sand Hill Road, Suite 280
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025-7022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,220
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/306,542
FILING DATE: 14-SEP-1994
APPLICATION NUMBER: US 08/160,965
FILING DATE: 02-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter Ph.D., Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: BAYL-004/01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6200
TELEFAX: (415) 854-3713
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes

STRAIN: MGAS 1719
IMMEDIATE SOURCE:
CLONE: SPB87 (cysteine protease)
PUBLICATION INFORMATION:
AUTHORS: Kapur, V.
AUTHORS: Topouzis, S.
AUTHORS: Majesky, M. W.
AUTHORS: Li, L.-L.
AUTHORS: Hamrick, M. R.
AUTHORS: Hamill, R. J.
AUTHORS: Patel, J. M.
AUTHORS: Musser, J. M.
TITLE: A conserved Streptococcus pyogenes
TITLE: extracellular cysteine protease cleaves
TITLE: fibronectin and degrades vitronectin
JOURNAL: Microb. Pathog.
VOLUME: 15
PAGES: 327-346
DATE: 1993

Harry Match	11.1%	Score 494;	DB 3;	Length 398;
Best Local Similarity	30.6%	Pred. No. 7.6e-31;		
Matches 129; Conservative	73;	Mismatches 135;	Indels 64;	Gaps 13;

```

Oy      6 LLAIYMEGLAMQG-----HSAPTYKERALSLARLALRQVSLRMGQTVASPKISIDYY 59
Db      10 LLSLALGFLVLANVPADQNFARNEKEKDSATIFIQSAAIKAGARBAED-ITLUDKN 68
Oy      60 RQGDAREGITSQEGSPAYFYVANGNNEGVALVLAADRIPITLAVSPYIGREDMSPDN 119
Db      69 LGSEL-----SCSNMYVYINISPG--GFIYVSODKRSPEILGTSGSTGDANG-KEN 116
Oy      120 LRMLQIT-----DQELIGLISGAQLNEELLRTREGVAEYHALMDNCHFANDPPKRNQ 173
Db      117 IASFESEVSEQIREKKKLDPTYAGAEKQPY-----AKSLUD-----SKGIHYNQ 162
Oy      174 GYPNNKEPLLPN-----GNAAVTGCVTAAAOIMRHYSHMPLQEGESFDY----- 218
Db      163 GNPYLLLPVIEKYKRGQSFVGQAAATGCVATATRQIKYINYPKMGKLDKDTYTLSSNN 222
Oy      219 ----HAGLVGCMGSGTFEGEMTDWIN-MPQNPDLNLUTOSQVDAVATLMDVSNASVMSFY 273
Db      223 PYFNHPKNI---FAAISTROYMWNHILPTYSGRESVQKM--AISELMDVGISVMDYG 277
Oy      274 ENGSGTSVYVYGALRNPNFRYKRSLOLHNRALYTSQEMDMTRGELASGRPYVYAGNNGS 333
Db      278 PSSGAGSRRORALKENFGYQSHQIKNRSFSSKQDMBAQIDKELSQNQPYTQGVGV 337
Oy      334 IGHAEVCDGVASDGTFFHFNMGVGSNGFYKJTLSPSLAGIGGEGIGFTIYOELLTGIE 393
Db      338 GGHAFVIGDADGRNFYHYVMWGMGVSDGFFRLDALNPALGIGGAGGFGNGVSAVYGIK 397
Oy      394 P 394
Db      398 P 398

```

RESULT 2
PCT-US95-11723-5
Sequence 5, Application PC/TUS9511723
GENERAL INFORMATION:
APPLICANT: Musser M.D., James M.
APPLICANT: Kapur M.D., Vivek
TITLE OF INVENTION: Methods and Compositions for
TITLE OF INVENTION: Identifying Streptococcus
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAW OFFICES OF BARBARA RAE-VENTER
STREET: P.O. Box 60039
CITY: Palo Alto
STATE: CA
COUNTRY: USA

```

ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11723
FILING DATE: 14-SEP-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/306,542
FILING DATE: 14-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Ventler Ph.D., Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: BAYL-004/0305
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6205
TELEFAX: (415) 424-8760
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
AMTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
STRAIN: MGAS 1719
IMMEDIATE SOURCE:
CLONE: SPEB7 (cysteine protease)
PUBLICATION INFORMATION:
AUTHORS: Kapur, V.
AUTHORS: Topouzis, S.
AUTHORS: Majesky, M. W.
AUTHORS: Li, L.-L.
AUTHORS: Hamrick, M. R.
AUTHORS: Hamill, R. J.
AUTHORS: Patel, J. M.
AUTHORS: Musser, J. M.
TITLE: A conserved Streptococcus pyogenes
TITLE: extracellular cysteine protease cleaves human
TITLE: fibronectin and degrades vitronectin
JOURNAL: Microb. Pathog.
VOLUME: 15
PAGES: 327-346
DATE: 1993
PCT-US95-11723-5

```

Query Match	11.1%	Score 494	DB 5	Length 398
Best Local Similarly	30.68%	Pred. No. 7.6e-31		
Matches 129; Conservative	73	Mismatches 155;	Indels 64;	Gaps 13

QY 6 LIAIIVLFGIAAOG-----HSAPTKREALSLIALAROYSLMGTAVDKSIDIYV 59
 Db 10 LLSLLALGGFVLANPVPADONFARREKEKKSADITFOKSAATIKAGARSAD -IKLDKVN 68
 QY 60 ROGDAERGITSOEESGPAYFYVANGNNEGVALVAADRIPTLIAYSPIGFHDMSPDN 119
 Db 69 LGSEL-----SSGNNMYVNIISTG--GPIYVSGDKRSPETLIGYSSGSFGAANG- KEN 116
 QY 120 LRMWLOIY-----DOEIGLILSGAQLNEELILRTGVPAYEVALMDNGHAFDPMRNQ 173
 Db 117 IASFMESEYEOIKENKIKLDTTYAGTAETKOPV-----VKSLD-----SKGIHYNO 162
 QY 174 GYPMNNKEPULLR-----GNAAVYGCVAATAAQMIRHSPLOGEGSFQV----- 218
 Db 163 GNPYNLLTFPIEIKVARGEGSPVGQAAAGCAVATAQIMKHNHPNGKGLDYDTYTLSSNN 222

QY	219	----	HAGSLVGWMCSTFGEMDYMIN-MPCNPDLNDLTQOQVAYALIMRVSASXSEY	273
Db	223	PFYHNPKUL----	FALISTROYIMNNILPPTSGRESVQKM--AISELMADVGISVDMQYG	277
QY	274	ENGSGTYSVYVVGALRNRFRRKRSQLHVRALYTSQEWMDMIGELASGRPVYVAGNMQS	333	
Db	278	PSSGAGSSRRVQBALKENRGVYNSHQJNRSPFSQDKWEQJDKELSSQNPVYVYGGVGKV	337	
QY	334	IGHAFVCCGIVASDGTFFHFRMGWGVSNCGYKATLLSPYLSLIGIGBEGITPTTIEITGIE	393	
Db	338	GGHAFVVIDGADGGRNFYHVNMGWGVSDGFFRIDALNPALSATGGGAGGFGNGYOSAVVGIK	397	
QY	394	P	394	
Db	398	P	398	
		P		

RESULT 3

US96-05997-1
Sequence 1, Application PC/TUS9605997
GENERAL INFORMATION:
APPLICANT: Musser M.D., James M.
APPLICANT: Kapur M.D., Vivek
APPLICANT: Ananthaswamy, H. N.
APPLICANT: Fernandez, A.
TITLE OF INVENTION: Use of extracellular cysteine protease
TITLE OF INVENTION: to inhibit cell proliferation
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: WEIL, GOTISHAL & NANCES
STREET: 2882 Sand Hill Road, Suite 280
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025-7022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05997
FILING DATE: 01-MAY-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/306,542
FILING DATE: 14-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/160,965
FILING DATE: 02-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter Ph.D., Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: BAYL-004/0205
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6200
TELEFAX: (415) 854-3713
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
STRAIN: MGAS 1719
IMMEDIATE SOURCE:
CLONE: speb7

Query Match	11.1%;	Score 494;	DB 5;	Length 398;
Best Local Similarity	30.6%;	Pred. No. 7.6e-31;		
Matches 129;	Conservative 73;	Mismatches 155;	Indels 64;	Gaps 13

```

OY      6 LTAIYMLFGIAMOG-----HSAPYTKERALSIALRALROVSLRMGQAVASDKISIDYU 59
Db      10 LLSLIALGFIYLANVYFADQNFARNEKEAMKDSAIFFIQKSAIKGARSABED-ITLKYV 68
OY      60 RQGDAMERGITSQEBSSPAYFYVYANFGNNEGVALVAADRIPTILAYSPIGRFMDMSDPN 119
Db      69 LGGEI-----SGSNMYVYVNIISG--GFIYVAGDKRSPEILYSTSSTGSDPANG-KEN 116
OY      120 LRMLLOIY-----DOEIGLLSGKAOLNEELTRFEGVPAEYHALMDNCHFANDPMRMO 173
Db      117 IASPFESTYEDIKENKKLDTTYAGAEIKOPY-----VKSLED-----SKGIHYNQ 162
OY      174 GYPNNKPELDPN-----GNHAYVGCVATAAQAQIMRHSWPLQEGESPDY----- 218
Db      163 GNPYNLLPVEIEKYVQESFVGQHAATGCVATATQAQIMKYHNYPNKGDKLDYTTLSNN 222
OY      219 ----HAGLVGNGMSTGEEMTDWIN-MGONPDLMDLQSOVDAYATLMDVASAYSMSY 273
Db      223 PYFNHPKUL---FAIISTRQYMNWNNILPFTYSGRESNVORM-AISELMAADVISYMDMG 277
OY      274 ENGSGTYSVYVYVYALRNFRYKRKSLQHLVRYALYTSQEWHDIMRGELASGRPYVYAGNNOS 333
Db      278 PSSSGAGSRRQYRALKEFNFGYVNSHOQINRSPFSQDNWMAQIDKLELSONQPYIYGVGV 337
OY      334 IGAHAFVCDGVASDGTFFHNMGMGVVNSGFYKJTLSPSLGIGGBCIGFTYIOEIIITGIE 393
Db      338 GGHAFVIDGADGRNFYHVNMGMGVGVSDEFFRLDALNPASALGTGGGAGGNGFYOSAVVGIK 397
OY      394 P 394
Db      398 P 398

```

RESULT 4

US-08-840-466A-19
Sequence 19, Application US/08840466A
Patent No. 6261561
GENERAL INFORMATION:
APPLICANT: Stewart, C. Neal
McKee, Marian L.
O'Brien, Alison D.
Wachtel, Marian R.
TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Alone Or As A Fusion Protein With One Or More Other
Antigens.
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finegan, Henderson, Parabow, Garrett &
Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/840,466A
FILING DATE: 18-Apr-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Boone, Laura S.
REGISTRATION NUMBER: 43,505
REFERENCE/DOCKET NUMBER: 04995,0029-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:
LENGTH: 934 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-840-466A-19Query Match 3.28; Score 140; DB 4; Length 934;
Best Local Similarity 17.4%; Pred. No. 0.023;
Matches 167; Conservative 128; Mismatches 331; Indels 336; Gaps 45;

```
74 GSPAYFYVNR--GNNEGVALVAADRIPTILAYSPIGRF-----DMSDMP-N 119
26 GLGLFFVYVONSPANGENEFKLGSDSKLTHDSYO--NRLFTYTLKGTGTADVLSKSDIN 83
120 L-RM---LQIYDOETIGLISGKAQINBEILRTGEPVPAEVALMDNGHFANDPARNNOGY 175
84 LSTVSLNKLHLYSSESEMKKAPGQ--QIILPLKLPREYSAL-----124
176 PNNKEPLPNC-----NHAATGCVATAAAQIMRHSMPLOG 212
125 PLLGSAFLVAAGVAGHINKLTKMSPDVTKSNMTDDKALNAAQAAASLGSQLSRSLNG 184
213 EGSPDYHAGSLVGNNGSTFGGEMTDINNPNPD-----DNLQSOYDAATILMDRVAS 267
165 DYAKTPTALG-IAGNQASS--OLOAWLQHYGAEVNLQSGNFDGSSLD-----229
268 VSMSEYEN-----GSGF-----YSYV-----284
230 FLPEFYDESEKMLAFQVGARYIDRFPAANLGAQREFLPANMLGYVNFIDQDESGDNTRL 289
285 -VGLRNRRYKRSLOLHVRLALTYSOEHDIMRGELASGRPVY-----YAGNNOISIG 335
290 GIGGEYWRDYFKSS-----VNGYFRMRMHESYHKRDYDERPANGFDIRNGILPSTPALG 345
336 HAFVCDGASDGTFFHFMGMGVSNGFYKLTLSPLTSLGIGEGGFTYIOBITTGIEPA 395
346 AKLYEQYGDVNALEFNSD-----KL--GSPAAATVGNVTPIPILVTGMI---389
396 KTPAAGTALPLALAKDIEAEYKSSGLNVGYSIYNTGEEQS--NLDLGRILNKADGE 452
390 -----DYR-----HGTGENNDLTXMQFRYQDKMSQ 417
453 VIEVKTSSINISWYGEHPESESFLAPNOLSGINTITLLYR-----TGTEQ 500
418 QIEPQ-----YVNELRLTSSGRDLYORNNITILEYKKODILSLNIPHDINGTEH 467
501 -----W-EPVHAAGGYVNSIKVNTTTPNNVVYVYVNDNEKSLIV 540
468 STOKIQLYKSKYIGDRIYWDSDALRSQGIQHSQSADQYAILPAYVQGS-----522
541 NSFVADLMSYHS-----TIVYONSDSPDEIRTPPAFAISGTATADVLSLGMW 591
523 NIYKVTARAYDRNGSSNNVOLITITVLSNGQYVDQVGTDTADKTSAKADNDADITTYTA 582
592 AEVPGG--SSNYPVVMSKDVLTLLSEGGYTLWYRESINNOKEDEKKIGSVKPTPEYTPR 649
583 TVKKNQVAQANVPVSPFN-----IYSGTATL-----GANSKATDANGKATYTLSSIP- 629
650 LEVGHNOTSTYTLDMANRNVLPDFLKNLGLPFNELVYVVRQTOSSSGSLMAAQEYVH 709
630 -----GQVYVSAKTAEMS-----ALNSAVIIFDQTKASITEI-KADKTTA 670
710 IKOGE--TFVYKRPVYBG-PIDGSTR-ATLAAVNGQOQLYL-----K 748
671 VAANGKAIKTYVKNMGQPVNNOSVTFSTNFGMFGKSGQQTQATGNGRATITLTSSSA 730
```

```
QY 749 GRNRYTVNGTAVEATEISS--EIRV-----FPNARDYVEISAPC--- 789
DB 731 GKATVSATVSDGAEVKAETVTFPELKDINKVDIIGNNNGELPNITLQYGFKLKASCG 790
QY 790 -----IPQETSIPLFDSKRYKWSLSAGHRMDVSRPLNG-----AYILKVGCIYTK 838
DB 791 DGTYSWTSENTSIATVDASGVTLN-----GKGSV-VIATSGDQYVSYTIKAPSYMIR 844
QY 839 IN 840
DB 845 VD 846
```

RESULT 5

US-09-696-188B-19
Sequence 19, Application US/09696188B
Patent No. 6406885

GENERAL INFORMATION:

APPLICANT: Stewart, C. Neal

O'Brien, Alison L.

Machiel, Marian R.

TITLE OF INVENTION: Method Of Stimulating An Immune Response

By Administration Of Host Organisms That Express Inflam
Antigens.

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:
ADDRESSEE: Finegan, Henderson, Farabow, Garrett &

STREET: 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/696,188B

FILING DATE: 26-Oct-2000

CLASSIFICATION: <Unknown>

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/840,466

FILING DATE: 1997-04-18

ATTORNEY/AGENT INFORMATION:
NAME: Boone, Laura L. S.

REGISTRATION NUMBER: 43,505

REFERENCE/DOCKET NUMBER: 04995, 0029-00000

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 19:

US-09-696-188B-19

Query Match

Best Local Similarity 3.28; Score 140; DB 4; Length 934;
Matches 167; Conservative 128; Mismatches 331; Indels 336; Gaps 45;

```
QY 74 GSPAYFYVNR--GNNEGVALVAADRIPTILAYSPIGRF-----DMSDMP-N 119
DB 26 GLGLFFVYVONSPANGENEFKLGSDSKLTHDSYO--NRLFTYTLKGTGTADVLSKSDIN 83
QY 120 L-RM---LQIYDOETIGLISGKAQINBEILRTGEPVPAEVALMDNGHFANDPARNNOGY 175
```

Db 84 LSTIMSLINKHYSESEEMMAEQ--QITLPLKLPFEYSAL----- 124
QY 176 PMNKEPILPNC-----NHATYGCATAAQIMRHSNPLQ 212
Db 125 PLIGASPLVAAGVAGHTNKLTKMSPDVTKSNMTDKALNTYAAQOASLSQSLSLNG 184
QY 213 EGSFDYHAGSLVNMGSGTEGEMTDMINPCNPDL-----DNLQSOVDAYATLMRDVSAS 267
Db 185 DYAKDTALG-IAGNOASS--QLOAWLQHGTAEVNLAQSGDNFEGSSLD----- 229
QY 268 VMSKFEYEN-----GSGT-----YSVV----- 284
Db 230 FLPEFYDSEKMLAFQGVARYIDSRFTANAGACREFLPANMLGVNFIQDPSGDNTRL 289
QY 285 -VCAIANNRRYKSLDLHRAKLTSEMDHIMGELASGRPVY-----YAGNNOSIG 335
Db 290 GIGGEYWRDYFKSS--VNGYFRMRMRHESYHKKDYDERPANGFDIREFNGYLPSPYALG 345
QY 336 HAFVCDGASDGTFFHNMWCGVSGNGFYKLTLSPTSLGIGEGIGFTIYQEIITGIEPA 395
Db 346 AKIYQIYGDVNLFNDS-----KL-----QSNPGATYGVNTPIPLTYMGI--- 389
QY 396 KTPAEAGTDLPLALKIEAEKSSGLNVGYSIYNGEES--NLDIGYRLNRADGE 452
Db 390 -----DTR-----HGTGNENDLTXSMQFRYQFDKSMQ 417
QY 453 VIEVKTSSINISMYGGEHPESLAPNLSGGINITTLXYR-----TGTG 500
Db 418 QIEPO-----VYNELRTLSGRYDLVORNNNILEKKODILSLNIPHDINGTEH 467
QY 501 -----W-EFVRHAGCGVNSIKVNTDPANNVVVVDNNEKSLIYV 540
Db 468 STOKIOLIVKSKYGLDRIMWDSALRSQGGQIOHSGSQSADYQALIPAYOGS----- 522
QY 541 NSFVADLNSTYHS-----TIVQFNSDPELRTPVAFALSTGATADVYSLGWM 591
Db 523 NIYKTAAYADRNNGSNNNVOLTITVLSNGOVVDQGVDTFADKTSAKADNDTIYTA 582
QY 592 AEPGCG--SSNPVYVSKDVLTLSEBDYTLMTYRFSINNOKDEKKIGSVSKPTETHP 649
Db 583 TYAKNGVAQANVVSFN-----IVSGTATL-----GANSATTDANGKATVILKSTP--- 629
QY 650 LFEVGHNOTSTYTLDAHNRVLPDFTLKNLGLPFNGELVYVFRQTOSSGSLAAOETVH 709
Db 630 ----GGVVVSAKAEHSS-----ALNMSAVIEFDOTAKAITEI-KADKTA 670
QY 710 IKOGF--TFVYKPVVEG-PIDGSYR-ATLHAFVNGQOQLYL-----K 748
Db 671 VANGKDAIKYTVVMKNGQPVNNOSVTFSTNFGMFNGKSQTOATTGNDGRATITLSSSA 730
QY 749 GKNTYVYKIVNGFAVEAIESS--EETRIV-----FPNARDVVEISAPC--- 789
Db 731 GRATYATVSDGAEVATEVTEFDELKIDNKVDIGNNVAGELPNTLQOGQEKRLASGG 790
QY 790 -----IPESTIILFDLSGKIYMKNSLSAGHGRMVSRLPNC-----AYILVDGYTK 838
Db 791 DGTYSYSENTSIATYDASGKVTLN-----GKGSV-VIATSDCKQTVSTITAKPSMIK 844
QY 839 IN 840
Db 845 VD 846

RESULT 6
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159
Query Match 3.2% Score 140; DB 4; Length 10182;
Best Local Similarity 18.8% Pred. No. 1;
Matches 149; Conservative 116; Mismatches 282; Indels 244; Gaps 42;
QY 148 TEGVPAEVALMDNGHEPANDPMRWNOGYPMNKEPILPNCNHATYGCATAAQIMRHS 207
Db 1819 SQSIPKQIYETTINGRP-----NSSGDAYPGNFQAVNQY----- 1853
QY 208 WP-----LQSGSPDYHAGSLVGNMGTGEMT-----DWINM-----PCNPDDLND-- 248
Db 1854 WPEHMDFRMAOGSGT--PSSRNAGSFTKTVTVYONGQTEYNNVLFKVPKPKPVIDSNS 1910
QY 249 -LQSOVDAYATLMRDVSASVMSFEY-NG-----SGTYSVYVVAL-RNNFR 293
Db 1911 VISKQOLNGOQILYRNVPQAOVTLVOSKTVLPNTYTIDSNGIATVTYQGLPLPGNIT 1970
QY 294 YKRSLOLHVALYTSOEMHDMIRGELASGRPVYAGNNO--SIGHAFVCDG----- 342
Db 1971 AKTSMNNV--TYTKON--SSGIASNTTEDISVSESDQVNTAGMOKNDGIKIKGTN 2027
QY 343 -----YASDGFHNMWCGVSGNGFYK-----LILSPSLGIGEGIGFTIYQ 386
Db 2028 YNENDFNFSINIPASHTLWNEEPSWKNNIGTTTKTVTLPLNMQGTRTVDIPITTY 2087
QY 387 EITGTGEPKTPAE-----AGTDALPIL-----ALKDIEAEKSEGLN- 425
Db 2088 TV-----TAKNPRDQGRMLTNGTGYNNIIEENNRRLGTSKMDNRQPKNIAGVN 2142
QY 426 ----VGYSIYNGEES--NLDIGYRLNRADGEVIEKTSINISMYGGEHPESF 475
Db 2143 LIALVNNPGISTPLEVNVKVVYNNFDTQIYK-----IQVDTFPKGTAGYKHLDEG 2197
QY 476 SLAP-----NQLSGITITLLYKRTGEQEPYRNAOGGVNSIKVNTDPNNVY 527
Db 2198 EGPLIDGKRYMNOQSTG-----TTSQMOQLAYTRTFPVKGTGYDVNPSMWGV 2247
QY 528 TVDNNECKLSIVNSFVADLNSEHSHITVQFNSDPELRTPVAFALSTGATADVYSL 587
Db 2248 WQTSQSAKF-IVNNA-----KRNQPIIT--QSKTDVYTP-----GAYRNLILIS- 2289
QY 588 GWMAEYVGGSSNYPVVMSKDVLTLSBGDTLMTYRESINNOKDEKKIGSVSKPTETHT 647
Db 2290 -----GTNYY-IOASADKIVINKNGKL--TFEVKANNDRW-----TVERGSP--- 2329
QY 648 HPLFEVGHNOTSTYTLDAHNRVLPDFTLKNLGLPFNGELVYVFRQTOSSGSLMAOET 707
Db 2330 -DINGIGPTNNGT-AISLSRLAVRPGSISALATGSGETI-----STSTATSEIY----- 2377
QY 708 VHIKOGETFVYKPVVEGPIDGSYRATLHAFVNGQOQLYLKGR-----NYTVK 756
Db 2378 -----YKAPQPE--OATHTYDNGTIDILPDNSRNSINPTEYEINTEK 2421
QY 757 YVNGTAVE--AIESSEIRVFPNPARDYVEISAPCIPOETSIILFDLSGKIYMKNSL-- 811
Db 2422 L-NGNENQKSFITTKNNNGKWTINNKNPNVYEFN-----QNGKVVVF--SANTIKPNSQIT 2473
QY 812 ----SAGHGRMD 819
Db 2474 ITPKAGGNT 2484

RESULT 7

US-08-750-532-9

Sequence 9, Application US/08750532
Patent No. 5736339

GENERAL INFORMATION:

APPLICANT: MITTA, Masanori

APPLICANT: YAMAMOTO, Katsuhiko

APPLICANT: MORISHITA, Mio

APPLICANT: ASADA, Kiyozo

APPLICANT: TSUNASAWA, Susumu

TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.

STREET: 419 Seventh Street N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: United States of America

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/750,532

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP95/01095

FILING DATE: 05-JUN-1995

APPLICATION NUMBER: JP 1994/130236

FILING DATE: 13-JUN-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 1994/173912

FILING DATE: 26-JUL-1994

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: MITTA-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528

INFORMATION FOR SEQ. ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1398 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-750-532-9

Query Match

Best Local Similarity 3.1%; Score 138; DB 1; Length 1398;

Matches 157; Conservative 91; Mismatches 262; Indels 296; Gaps 41;

DB 185 PNCNHAATGCVATAAQAQIMRYHNSPLOGESFDDHAGSLVGNMGTGEMWIMGNP 244

DB 351 PNCGEAVRG-----WDGHHGT--HVAAGTVAGYDSN--NDAMDWLSMYSGE 392

DB 245 -----DLNLTQ-----SQVDAYATIMRDVSAS----- 267

DB 393 WEVFSRLYGMVDTVTDTVGVAQAQIMALIRVLRSDGSGSMMDIEGNTYAATGADV 452

DB 268 VSMSEF-----YENSGRTSYVY-----VG 286

DB 453 ISMGLGNAPYLDGTDESVAAVELTEKYGVFVIAAGNPGINITYSGVATKAITYG 512

DB 287 ALRNNFYKRSIQLHVALYTSQEMHDMINGELASGRPVY-----AGNDSIGHAFFVC 340

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

513 A-----AAVPLNV-GYVVSQ-----ALGYPDYGGFYPPATVNR/IAF-FSS 552

341 DGVASDGTFFHFNW--GMGVSNGEYFKLLTSPSTIGIGEGGFTTYEITGIEPAKT 397

553 RGRPRIDEIKPNVAPGYG-----YISLPMWIG-----ADFGSGTSM-T 593

398 PAKGDAALPILAKIDIEAEYS-----ESGLNVGSIYVNGEESNLIDGVR----- 445

594 PHSVGVALLISGAKEGIYVNDILIKKYLESGATVLEDDPYTGKYTELDGGRVAVT 653

446 -----LNKADGEIEYKTSINISWYGEHEPESFAPNOISGINTITLLYRTGTE- 499

654 KSWELIKALNGTTLPIYDMADKSYSDAEY-----LGVDVIRGLYANSTPD 701

500 --QMEVRYAAGGYVNSIVNTTDP-----NNVYVVDNNEGKLSIVPNSFVADLNSYE 551

702 IYEMH-IKYVGDTEYKTEIYATEPMIKPFVGSVILENTEFVLRV-----KYDGELE 755

552 HSTIYVQFNSDSPDEIRPVPVAFALSTGATADVLSGVVMAEVRGSSNYPVMSKDVLT 611

756 PGLYVGRITIDP-----TTPV-----IEDEL-----NTIYIEK---F 787

612 LSEGYTL-WRFPSINNOKDEWKKIGSVYKTPLEYTHPLFEVGHNOTSTYLLMANNRV 670

788 TPENNYTILTWY-----DINGPEVTHHFTTYEGVDYLAMTYYDYG 830

671 L--PD-----FTLKNL-----GLPFNGELVYV---FROTOS-----SGSLW 702

831 LYRPDMGVFPYQDLYLAANSNPPGMEVMTGFENFAPLYEGFLVRYGYEITPSVW 890

703 AAQET-----VHIKQETEVYKPYVEG--PIPDGSYRATLHAIVNGQOOLKGRNY 753

891 YINRTYLDNTNEFSIEFNTINIVAPINATLIPILGYNMSVESV--GDEFFIKG--- 944

754 TKIYNGTVAEPAISSEIRFPPARDYVEISAPCI-PQETSIILFDLSGKIVMNSIS 812

945 -----IEVPGTIELK-----IRGNPSVNSDLIDLYLDSKGNLVALDGNP 986

813 AGHRMDVSRIPNGAVYILKYDGYTTR 838

987 TAEEVVEYKPKPGVSIYVGVSVR 1012

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

PRIOR APPLICATION DATA :
 APPLICATION NUMBER: PCT/JP96/03253
 FILING DATE: 07-NOV-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 323285/1995
 FILING DATE: 12-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Browdy, Roger L.
 REGISTRATION NUMBER: 25,618
 REFERENCE/DOCKET NUMBER: TAKURA-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 628-5197
 TELEFAX: (202) 737-3528
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1398 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 8-894-818b-8

[illegible]

OY 754 TVKIYNGTAVEAIEESSEIRVPPNPARDYVEIASACIP-OE5IILFDLGRIVKMSLS 812
 Db 945 -----IEVPETABLEK-----IRIGNSVPSNDDLYIDSKGNLYALDGNP 986
 OY 813 AGHGMDVSRLEPNCAYILKVDGYTK 838
 Db 987 TAEDEVVEYPRKPGYSIVVHGYSVR 1012
 RESULT 9
 US-09-445-472-6
 : Sequence 6, Application US/09445472
 : Patent No. 6358726
 : GENERAL INFORMATION:
 : APPLICANT: TAKAKURA, HIKARU
 : APPLICANT: MORISHITA, MIO
 : APPLICANT: SHIMOJO, TOMOKO
 : APPLICANT: ASADA, KIYOZO
 : APPLICANT: KATO, IKUNOSHIN
 : TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 : FILE REFERENCE: TAKAKURA-6
 : CURRENT APPLICATION NUMBER: US/09/445,472
 : CURRENT FILING DATE: 1999-12-06
 : PRIOR APPLICATION NUMBER: 151969/1997
 : PRIOR FILING DATE: 1997-06-10
 : NUMBER OF SEQ ID NOS: 33
 : SOFTWARE: PatentIn version 3.0
 : SEQ ID NO 6
 : LENGTH: 1398
 : TYPE: PRT
 : ORGANISM: Pyrococcus furiosus
 : US-09-445-472-6

[illegible]

Db 719 TGSVEVTAKTGDIKGIIESNSGVNITASGDTLNV 754

RESULT 12

US-09-206-942-34
Sequence 34, Application US/09206942
Patent No. 6432669
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae H1gh
TITLE OF INVENTION: Molecular Weight Proteins
FILE REFERENCE: 1038-861 M1S:Jb
CURRENT APPLICATION NUMBER: US/09/206,942
EARLIER FILING DATE: 1998-12-08
PRIOR APPLICATION NUMBER: 09/167,568
EARLIER FILING DATE: 1998-10-07
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 34
LENGTH: 1228
TYPE: PR
ORGANISM: Haemophilus influenzae
US-09-206-942-34

Query Match 2.9%; Score 128.5; DB 4; Length 1228;
Best Local Similarity 19.8%; Pred. No. 0.29;
Matches 102; Conservative 79; Mismatches 220; Indels 115; Gaps 20;

QY 129 OEIGLL-----SGKAOLNEILTEGVAPEVH-----LMNGHFANDPMRNOG 174
DB 329 QYTGVEIESONSASGSSIK---FKSEG---STHAFITKNDLILMANGGISTLNOVAG 382
QY 175 YPMNNEPPLLPGNHAYTCVATAAQMRYHSMPLQEGSPDYHAGSLVNGSGTEGEM 234
DB 383 ISNKLKSLIANKNITFECCGNTLAD-----KKPIKGNITVKEGANTLRSANYGND 437
QY 235 YDMINRPN-PDLNNTOS-----QVDAVATLMDVSVASVMS--FYENGSGTYS 281
DB 438 KSAIIRGVNTKNGITVGSALINIEKNITVEGSAFELANPNYSFVNSGLFDQGSNIS 497
QY 282 VYVVGALRNFRKRSIQALVRYLTQSEWDMIRGELASGRPVYVAGNNSIGHAFVCD 341
DB 498 IAKGAIIFKDIETGSLNITTK---SDSNHHTIIRKGNITRKCDLITNN-----544
QY 342 GYASDGFPHNMWGVNSNGFYKLLLSPTSLGIGEGIGFTTYOELITGIEPAKTPAEA 401
DB 545 ---GDWT-ELQIG-GNISQKEGNTL-----ISSDKVNIITERITIKAGVNGDNDSDNE 591
QY 402 GTDALPILAKDIEAYKSESGLNVGYSTYN---TGEOSNLDLGRLLKADGEVTEYK 457
DB 592 ASANLITTK---ELKLTNDLNT---SGFNKAETIAKDNSNLTIDNSAGWTDAKKY 645
QY 458 TSSINISWYGYGHEPESFLAPNOLSGITITLRLRTGEQMEVVRHAGGVASIKY 517
DB 646 FSNVKS-----KISASHNVTL-----NS-KV 667
QY 518 NTDPPNNVYVYDNNKSLIYVNSFVADLNSYHSTIYVQFNSDPELIRTPVAPALST 577
DB 668 ETSGDPTSTEDGGNNNTGLITPAKNVTYNNNITLSHKTIVNTASENTTKAGTTIN---AT 724
QY 578 GATADVDYISLGMVAEVPBGSSNYPVWMSKDYTLIS 613
DB 725 TGSVEVTAKTGDIKGIIESNSGVNITASGDTLNV 760

RESULT 13
US-08-624-655A-2
Sequence 2, Application US/08624655A
Patent No. 6323005
GENERAL INFORMATION:

APPLICANT: DABAN, MONTERRAT
APPLICANT: MEDRANO, ANDRES
APPLICANT: ESPUNA, ENRIC
APPLICANT: QUEROL, ENRIQUE
TITLE OF INVENTION: TRANSFERIN-BINDING PROTEIN 1 (TBP1) GENE OF
TITLE OF INVENTION: Actinobacillus pleuropneumoniae, ITS USE TO PREPARE
TITLE OF INVENTION: PRODUCTS FOR THE UTILIZATION IN VACCINES FOR
TITLE OF INVENTION: PLEUROPERNEUMONIA AND AS DIAGNOSTIC REAGENTS
FILE REFERENCE: P00740
CURRENT APPLICATION NUMBER: US/08/624,655A
CURRENT FILING DATE: 1996-03-22
PRIOR APPLICATION NUMBER: 95 00592
PRIOR FILING DATE: 1995-03-24
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 931
TYPE: PR
ORGANISM: Actinobacillus pleuropneumoniae
US-08-624-655A-2

Query Match 2.9%; Score 128; DB 4; Length 931;
Best Local Similarity 19.3%; Pred. No. 0.2;
Matches 146; Conservative 104; Mismatches 286; Indels 222; Gaps 43;

QY 57 YVTRQG-----DAE---RGIT-----SOEGSPATFYVANNRNNEGYALVAADRIPTI 102
DB 208 YTHRDKETRAHKAHDAESRSONITRVGETNELDTSNRYTTNNQHTYGFILKDEPPL 267
QY 103 -LAYSPIGRFMD-----SMDNLRMWQIYDOEIGLISGAOLNEILRTGVAPEVH 156
DB 268 DCFPKAMARTKTPSPRSYPTETPREKQAYENQKH-----TERNADYGEYR 318
QY 157 ALMDNGHFANDPMRNOGYPMNKEPILPNGNHA-----YTCVATAAQMIR---204
DB 319 ALDPLKXKSDSLVLYGTFSEKHYVAGYEHSEKORYDRTDYTAAYMOPSLRTGRN 378
QY 205 -----YHSMPLQEGSPDYHA-----GSLVGMN--SGTFGEYMDWI 238
DB 379 WYPMNNAKGLYRDNALDGV-AIDFTEDGYKSSKGLMMAKARIDEMHTDRGLALRYT 437
QY 239 NMDGNPDLNLT-----QSOVDAYATLMDRVASV-----SMSFE 274
DB 438 NQGNRLIDRLSLSPDQKINLSTRLENNCSEPTIDKNCRAITDLKMSSTKNEGSYE 497
QY 275 NGSCTYSV-----YVVGALRNFRKRSIQALH-VRLYTSQEHMDIRGELAS 321
DB 498 EKHDTIQLSDKTIVOTGLGKHQNLMLGSDRFNSTLKRHEILSEFSVGTWH-RIRNGYK 556
QY 322 GRPVYVAGNNSIGHAFVCD-----G 342
DB 557 DTPYITELKQALYISKNECYSGTIGACADCATSKIGHNHYIALRDNFAITYLIDIGL 616
QY 343 YASD-STFHRNMW---GGVNSGFYKLLTLLS-PTS-LGI---GEGIGFTYOEII---389
DB 617 YRDKDKFRSTHMANAGDVKNSAMNIGIYAKTSPSLSTRASSGGRVSPFELGLRY 676
QY 390 TGIETPAKTPAAGTDLAL-PILAL-KDIEAYKSESGLNVGYSTYNMGEOSNLDLGRN 447
DB 677 DGAHSSSDAYQKTEKISPEKSLNOEVAALFFKDFGV-VEVSFYKNDYKOL-TAPERMH 724
QY 448 KAGG-----EIVKXSSIN-----ISWYGEHPESSFLAPNOLSGOINITLRYRTGT 498
DB 735 QIOSMTNIFNVQIKLDGILIKLDMNGVFD-----KIPGITY-TLAYSMARY 783
QY 499 EOWEPYRHAOGGYVN-SIKVNTDPPNNVVVTV--DNNKSLIYVNSFVADLNSYHST 554
DB 784 KE--VKNYQ-GYMNIRSPDLDTIGPARVYVGVGQDPDEKMGV-----N 824
QY 555 ITVOENS-DSPDEIR--TPYAFALSTGATADVDYISLGMVAEVPBGSSNYPVWMSKDYLT 611
DB 825 LFTWTHSSGKNPNELRGNNEGVGFA-NYERTATKRTKRLTSMHTFDLTG-----YTPWKH--TT 877

QY 612 LSEGDYTLV-YRFSINNOKDEMKKIGSVKPTPEYTH 648
DB 878 VRAGVYNLMNRYT-----TWESYROSSLAINHOTH 909

RESULT 14

US-08-913-942-15

Sequence 15, Application US/08913942
Patent No. 6200578

GENERAL INFORMATION:

APPLICANT: St. Gene, Joseph

APPLICANT: Barenkamp, Stephen J.

TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/913,942

FILING DATE: 29-DEC-1997

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,995

FILING DATE: 24-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/4031

FILING DATE: 22-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Vance, Dolly A.

REGISTRATION NUMBER: 39,054

REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAY

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 679 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-913-942-15

Query Match

Best Local Similarity 19.5%; Score 127.5; DB 4; Length 679;

Matches 138; Conservative 89; Mismatches 239; Indels 241; Gaps 33;

QY 242 GNPDLNLT-TOSQVDAYATLMRDVSAVSMSFEYENGSGTYSYVVGALRNFRKRSIOL 300
DB 86 GTTEVINLNTDSSGNAVGSSTITFKAGNLKIKOSG-----NDFTY--SLK 130
QY 301 HVRALYTSQEMHDMINGELASGRPRVYAGNNSIGHAFVCDGYASDGTFFHFMGCVSN 360
DB 131 ELKNL-TSVETEKLSFG--ANGNKVDITSDANGKLAKTGNGNGNSVHLN---GIAS 183
QY 361 -----GEYKL-----TLSPISLIGGEGIGFTYQETITGIEP 394
DB 184 TLTLTLAGGTGHTVDINDAVNHRAASVQDVLNSGNNIGNGNNDVFRYDVTVDYVNG 243
QY 395 A-----KTPAEGTDALPILALKDIEAEKSESQ--LNVGSIYNTGSEOS 438
DB 244 ANANVSATADTAHKKTTVRVDVYGLPV-----QVTEDEGKTVYKVGNEYK-AKDDG 294

QY 439 NLDLGRNLKADGEVIEVTSINISIMYGGEHPESFSLAPNLSOGINTITLLYRRGT 498
DB 295 SADMNQKVE--NELKRTVKLVAS-----GTPVRSINAVDGT 332

QY 499 EOMEPEV-----RHAOGGVN-SIKVNTDPNNVYTVVNNKGLSI 538
DB 333 EDTDAVSFNOLKALODKQVTLSTNSNAYANGTDNDGKATQTLNSGLNFKFSSGDELK 392

QY 539 VPRSFAVDINSYEHSTITVQFNSDSDPELRTVVALSLSGAT-----ADVISL 587
DB 393 I--SATGDVITTPPKGSIVQVGGDKAST-----SKGANTTEGLVEASELSLKL 442

QY 588 GW-VMAEVPG-----GSSNYPVWSKDYTLTSEGD-----YTLVYRFSINNOK 629
DB 443 GKVGVGEKVGSEGLDTSKETLVKSGDKVTYLAGNKLKVGQEGTFTAL-----K 493

QY 630 DEMKKIGSVSVTPPEYTHPLFEVGHNOTSTYTLDMANRVLDPFTLLKGLPENGELV 689
DB 494 DELTGKVSVEFDYAN-----GANAST-----KITKGLTITLANGANGATVT 537

QY 690 VERQTOSSGSLMAAQETVHIKQGETFYKPYEGPIPDGSYRATLHAFVNGQOOLV--- 746
DB 538 DADKIKVASDGISAGNKAV-----KNVAAGEI-----SATSDAING-SQLAVA 581

QY 747 -----LKKRNYTVKIVGTAVEAIESEELRVPNPARDVETISACIPQET 794
DB 582 KQVTLACQVNNLEKGVKVRADAGTASALASQ-----LPQAT 622

QY 795 SLIFDLGKIVYKNSLSAGHR-----MDVSRLL-PNGAYITLVQDGYT 836
DB 623 -----MPGKSMVSIAGSSYOGQNGLAIGVRSIDKNGKVIIRLSGTT 663

RESULT 15

US-09-268-347-26

Sequence 26, Application US/09268347

Patent No. 6335182

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS

FILE REFERENCE: 1038-860

CURRENT APPLICATION NUMBER: US/09/268,347

CURRENT FILING DATE: 1999-03-16

NUMBER OF SEQ ID NOS: 54

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 26

LENGTH: 679

TYPE: PRP

ORGANISM: Haemophilus influenzae

US-09-268-347-26

Query Match

Best Local Similarity 19.5%; Score 127.5; DB 4; Length 679;

Matches 138; Conservative 89; Mismatches 239; Indels 241; Gaps 33;

QY 242 GNPDLNLT-TOSQVDAYATLMRDVSAVSMSFEYENGSGTYSYVVGALRNFRKRSIOL 300
DB 86 GTTEVINLNTDSSGNAVGSSTITFKAGNLKIKOSG-----NDFTY--SLK 130
QY 301 HVRALYTSQEMHDMINGELASGRPRVYAGNNSIGHAFVCDGYASDGTFFHFMGCVSN 360
DB 131 ELKNL-TSVETEKLSFG--ANGNKVDITSDANGKLAKTGNGNGNSVHLN---GIAS 183
QY 361 -----GEYKL-----TLSPISLIGGEGIGFTYQETITGIEP 394
DB 184 TLTLTLAGGTGHTVDINDAVNHRAASVQDVLNSGNNIGNGNNDVFRYDVTVDYVNG 243
QY 395 A-----KTPAEGTDALPILALKDIEAEKSESQ--LNVGSIYNTGSEOS 438
DB 244 ANANVSATADTAHKKTTVRVDVYGLPV-----QVTEDEGKTVYKVGNEYK-AKDDG 294
QY 439 NDLGVRNLKADGEVIEVTSINISIMYGGEHPESFSLAPNLSOGINTITLLYRRGT 498

Db 295 SADNOKVE--NGELAKTKVLSAS-----GTNPVKISNADGT 332
QY 499 EOWEPV-----RHAOGGYV-SIKVTPDPNNVTVVDNNEGLST 538
Db 333 EBDNVSFKQKALQKQVLTSTSMAYANGTNDGKATQTLISNLFKSSDGLK 392
QY 539 VPNSFVADLINSYESTITVOFNSDSPDEIRTPVAFALSTGAT-----ADYISL 587
Db 393 I--SANGDYTFPPKGSVOVGDDGKASI-----SKGANTTEGLVEASELVEISNKL 442
QY 588 GW-VMAEVRG-----GSSNPVWVSKDVLTLSEGD-----YTLWYRFSINNOK 629
Db 443 GMYVEKVGSGELDSTKETTILVSGDKVTLKAGDMLKVQEGSTNFTYAL-----K 493
QY 630 DEKKKIGSVYKPTETHPLEFGVHNGSTYTLDAHNRVLPDFTLKNLGLFPNGELV 669
Db 494 DELTGKSVYEFKPTAN-----GANGAST-----KITKDGTLITLANGANGATV 537
QY 690 VFRQTOSSGSLMAAETVHIKQETFEVYKPYVEGPIDGSYRATLHAFVNGOOLY--- 746
Db 538 DADKIVADGIGAGKAV-----KNVAGEL-----SATSDAING-SQLYAVA 581
QY 747 -----LKGRNTYVKIYNGTAVEAIESSEIRVFPNPARDYVEISAPCIPOET 794
Db 582 KGVNLAGOVNNEGKYNKVKRADAGTASALAAQ-----LPQAT 622
QY 795 SILFPLSGKIYVKNLSAGHR-----MOYSL-PNGATYILKVDGT 836
Db 623 -----MPGKSWSIAGSSYOGNGLAIGVSRISDNKVIIRLSGT 663

RESULT 16

US-08-277-231A-4
Sequence 4, Application US/08277231A
Patent No. 5643725

GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pillin
Patent No. 5643725
TITLE OF INVENTION: Structural Genes and The LKP Pill Operon of No. 5643725cytable
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,231A
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 32,542
REFERENCE/DOCKET NUMBER: ACC94-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 741 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-277-231A-4

Query Match 2.88; Score 125; DB 1; Length 741;
Best Local Similarity 21.18; Pred. No. 0.24;
Matches 128; Conservative 76; Mismatches 216; Indels 188; Gaps 32;

QY 22 APVTKERALSIALRLR-----QVSLRQGTAVSPKSIDYVROGDAERGITSGEG 74
Db 186 APVYRGANNTMAVSTIKONQYITQITVPAGPVIINDLVASG---SGDLTVEI-QESDG 241
QY 75 SPAYFYVANGNNEGVALVADRIPITLAYSPIGRFDMSPDNLRLMLQIYDQEI- 132
Db 242 KVSFIVP-----FSNLAPLRVGLHFLQAGGRYRIDS-----RTFDERVLOG 285
QY 133 LI--LSGRKQMLEILRTGCVPAEVAHALDNHFAADPM-RNNQGVPMNKE-PILPNG 187
Db 286 VIQYGLTNLTLNLSLTYTHRYAGLF---GFLNTPIGAFSADATWSHAEPFLKHS 340
QY 188 NHAY-----TCGVAATAAQIMRYHSMPL----- 210
Db 341 KNGYSLHGSYSINPNSGTYITLAA--YRISSDFTYLTDTIGLNTFRPFGSAYLPEY 398
QY 211 QGEGSFEDYHAGSLVGNKSGTF--GEMYDWINPQNPDLNLTOQVDAVATLMDVASY 268
Db 399 RPKNQFVSLQSLGKMGNLYLSGQTYNWEKRGTN-----TOYOV-AYSNSFHLNYSV 452
QY 269 SMS--FYENGSGTYSYVVGAL-----RNNFRYKS-----LOLHYRLYTSQEMHDM 314
Db 453 NLSQSIDKETGKDNSTYLSLPLDGNHSADSSYSNGDINORLGVNSF--GERHQM 510
QY 315 IRGELASGRPVYVYAGNQSIGHAFVCDGYAS-DGFHFHNGMGVSNGFYKLTLSPTSL 373
Db 511 SYGIMS-----RNNQ-----GYSTYGNLSHNNSISY--RASYSROSLKRSI 553
QY 374 GIGEGE-----GFTYQETITGIEPAKTPAEDGADLPILALKDIEAKESGLANG 427
Db 554 SLGASGAVVAHKRGITLSQPV-----GESFAIHAKD-AAGAKVESGANGVS 598
QY 428 YSIYNTGEBQSNLDLGRLUKADGEVLEVKTSSINISWYV-----GEHPSPSLAPQLS 483
Db 599 LDYF-----GNNAVPTSPREINVTICINSDAEANYFEATERQIT 639
QY 484 QGINTTILYRRGTGEQWEPVRAOGGYVNSIKVNTDPNNVVY---TYDNNEGKLSIV 539
Db 640 PRANSISLVDFRGNK-----TMVLENTLPLNGEPVPMASATQDSDG----- 681
QY 540 PMSFVADL 547
Db 682 --AFVGDV 687

RESULT 17

US-08-473-750-7
Sequence 7, Application US/08473750
Patent No. 5834187

Patent No. 5834187
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Jr., Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pillin
Patent No. 5834187
TITLE OF INVENTION: Structural Gene and the LKP Pill Operon of No. 5834187 5786
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,750
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,321
FILING DATE: 19-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC94-028
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 741 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-473-750-7

```

```

Query Match      2.8%; Score 125; DB 2; Length 741;
Best Local Similarity 21.1%; Pred. No. 0.24;
Matches 128; Conservative 76; Mismatches 216; Indels 188; Gaps 32;

```

```

OY 22 APVTKERALSLARLAR-----QVSLRMGQTAVIDSKISIDYVYRQGDAREGITSQEG 74
DB 186 APVARGVANTAKVSKINGNGTYITQITVPAGPVIINDLYASG---SGDLTVEL-QESDG 241
OY 75 SPATFYVANKNGNEGIALVAADRIPTITLAYSPIGRFDMSPMDLRLMLQIYDOEI--G 132
DB 242 KVRSEIYV-----FENLAPLMKRVGHLRYQLAGGRYRIDS-----RTFDERVLOG 285
OY 133 LI---LSGKAQLENEILRTGEPVPAEVAHMDNGHFANDPM--RMNOCYPMNKE--PLPNG 187
DB 286 VLOVGLTNHLTLNLSLITRYRAGLF-----GGLNTPIGAFSADATWSHAEPFLKHY 340
OY 188 NHAY-----TGCVATAAQMRYHSMPL----- 210
DB 341 KNGYSLHGYSINFNESGNTITLAA--YRYSRDFTLSDTIGLNRTPROFSGAILEY 398
OY 211 QEGSFYDYGSLVGNMSTF--GEMYDMINPMPDLDNLTSQVDAVATLMRDVASY 268
DB 399 RPKNQFOVSLSSLSLNGMGLYLSGQTYNWEKRGTN---TQYOV-AVSNSPHILNYSV 452
OY 269 SMS---FYENGSGTYSVYVVGAL-----RNNFRYKRS-----LQHYRALYTSQEMHD 314
DB 453 NLSQSIDKETGRKDSIYLSLPLGDHNSADSSYSRSGNDINOLGVNGSF--GERHOW 510
OY 315 IRGELASGRPYVYAGNNSIGHAFVCDGYAS--DGFHFRNMGVSGNGFYKLTLLSPTS 373
DB 511 SYGINAS-----RNNQ-----GYRSYDGNLSHNSISGY--RASYSRDSLKNSI 553
OY 374 GIGEGCI-----GFTIYOETITGIEPAKTPAEAGTDALPLIALKDIAEKSEGLAVG 427
DB 554 SLGASGANVAHKHGTLSQPV-----GESFALIHAKD--AAGAKVSGANVS 598
OY 428 YSIYNTGEOQNLGLYRLNKADEVELEVKTSSINISWYG-----GEHPESFLAPNOLS 483
DB 599 LDYF-----GNAVMPYSPYEINIGINIPSDALEANVEFEATEROIT 639
OY 484 QGINTITLLYRRTGEOMEPEVRAHAGGVNSIKYNTDPPNNVV---TYDNNESGLSIV 539
DB 640 PRANSISLVPERTKN-----TWLFNLTLPNCEPVPMASTADDSG----- 681
OY 540 PMSFVADL 547
DB 682 --AFVGDV 687

```

RESULT 18

US-08-477-326-7

Sequence 7, Application US/08477326

Patent No. 5968769

GENERAL INFORMATION:

APPLICANT: Green, Bruce A.

TITLE OF INVENTION: Sequence and Analysis of LKP P11n

Patent No. 5968769

TITLE OF INVENTION: Structural Gene and the LKP P11n Operon of No. 5968769typab

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

STREET: Two Millitia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,326

FILING DATE: 07-JUN-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/277,231

FILING DATE: July 19, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Carroll, Alice O.

REGISTRATION NUMBER: 33,542

REFERENCE/DOCKET NUMBER: ACC94-02A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 741 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-477-326-7

```

Query Match      2.8%; Score 125; DB 2; Length 741;
Best Local Similarity 21.1%; Pred. No. 0.24;
Matches 128; Conservative 76; Mismatches 216; Indels 188; Gaps 32;

```

```

OY 22 APVTKERALSLARLAR-----QVSLRMGQTAVIDSKISIDYVYRQGDAREGITSQEG 74
DB 186 APVARGVANTAKVSKINGNGTYITQITVPAGPVIINDLYASG---SGDLTVEL-QESDG 241
OY 75 SPATFYVANKNGNEGIALVAADRIPTITLAYSPIGRFDMSPMDLRLMLQIYDOEI--G 132
DB 242 KVRSEIYV-----FENLAPLMKRVGHLRYQLAGGRYRIDS-----RTFDERVLOG 285
OY 133 LI---LSGKAQLENEILRTGEPVPAEVAHMDNGHFANDPM--RMNOCYPMNKE--PLPNG 187
DB 286 VLOVGLTNHLTLNLSLITRYRAGLF-----GGLNTPIGAFSADATWSHAEPFLKHY 340
OY 188 NHAY-----TGCVATAAQMRYHSMPL----- 210
DB 341 KNGYSLHGYSINFNESGNTITLAA--YRYSRDFTLSDTIGLNRTPROFSGAILEY 398
OY 211 QEGSFYDYGSLVGNMSTF--GEMYDMINPMPDLDNLTSQVDAVATLMRDVASY 268
DB 399 RPKNQFOVSLSSLSLNGMGLYLSGQTYNWEKRGTN---TQYOV-AVSNSPHILNYSV 452
OY 269 SMS---FYENGSGTYSVYVVGAL-----RNNFRYKRS-----LQHYRALYTSQEMHD 314

```

Db 453 NLSQSIDKRTGKRDNSIYLSLPLGDNHSDSSYSGNDINQRLGVNGSF--GERHGW 510
 QY 315 INEGLASGRPVYAGNNGSIGHAFCVDDGYAS-DGTFHFNMGKGVSGFKLLSTSL 373
 Db 511 STYINMS-----RNNQ-----GYRSTYDGNLSHNSISGY-RASYSRSLKNRSI 553
 QY 374 GIGEGE-----GFTYOEIITGIPAKTPAAGTDALPILAKDIEAEKESGLNVG 427
 Db 554 SUGASGAVYAHKIGITLSQPV-----GESFALIHAKD-MAGAKYESGANYV 598
 QY 428 YSTYNTGEOSNLDLGRINKADGEVILEVKTSSINISWYGY---GEHPESFSLAPNOLS 483
 Db 599 LDYF-----GNAMVPTSPYEINVIYIGINPSDAEANYEFATERQII 639
 QY 484 OGINTITLLYKRTGTEOMEVPRAGGYNSIKVNTDPPNNVY---TYDANNEGKLSIV 539
 Db 640 PRANSISLDFERTGKN-----TMVLNLTLPNGEPPVMASTADSEG----- 681
 540 PMSFVADL 547
 682 --AFVGDV 687

RESULT 19
 US-08-620-717A-9
 ; Sequence 9, Application US/08620717A
 ; Patent No. 5670365

GENERAL INFORMATION:

APPLICANT: Fellelson, Jerald S.

TITLE OF INVENTION: Identification of, and Uses For, Nematocidal

NUMBER OF INVENTION: Bacillus thuringiensis Genes, Toxins, and Isolates

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saliwanchik & Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: Florida

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/620,717A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/540,104

FILING DATE: 06-OCT-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: MA94.C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (352) 375-8100

TELEFAX: (352) 372-5800

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1168 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: 167p

US-08-620-717A-9

Query Match 2.88; Score 125; DB 1; Length 1168;
 Best Local Similarity 18.58; Pred. No. 0.5;
 Matches 181; Conservative 147; Mismatches 368; Indels 282; Gaps 46;

QY 30 LSLARLARQVSLRMGOFAYSDKIS--IDVYRQGAERGITSOEESPAYVYVANGNN 87
 Db 70 LALTKASISLIGLPGDAVAPPTIMFVDFEPR--LFRGSGQDNRQAQFELLIEKVK 126
 QY 88 EGYALVAADR---IPITLAVSPIGRFDMSPMDNLMMQIYDOEIGLLSKAQOLNEE 144
 Db 127 E--LVDEDEFRNTLNINLVY-----LDGMQIALSHFQ--DYQALICGSGQPGMLMD 174
 QY 145 ILRTEGVPAVHALMDN-----GHPANDPMKMOGYPMNKEPL--LPNG 187
 Db 175 QPTFACPTPTDHLISVRESFKDARTTLETALPHKKNMLSTNNTPOFNSDTVLLTLP-- 232
 QY 188 NHAIVGCVATAAQAQIMKHYSPLOGBESFYIHAGSLVGNNSGTFGEYMDINMGNDLD 247
 Db 233 --MT-----TAATLNLHOGYIO-----FAERKSVNY--DES 263
 QY 248 NLTQSVDAVATATMRDVSASVMSFE---NCSGYVYVYGALRNF--RYKRSIQH 301
 Db 264 FINGTKVDLQRI--QDYSTVSTTFEKFRTLPNLSNKEV-----NKYNRYVRSMTL- 314
 QY 302 VRALYTSQEWHD-----INGELASGRPVY-----YAGNNOSIGHAFVCDGYASDG 347
 Db 315 -QSIDIATATPTLDNNVYPSNVDIQLDQTRLVFSDVAGPEGNDNITSNI----- 363
 QY 348 TFPHNMGWGVNGFYKLTLLSPSLGIGEGIGFTIYOELITGIEPAKTPAAGTDALP 407
 Db 364 -----IDVLTPIYNGIG-----FOE-----SSDKRTYTP 388
 QY 408 ILALKDIE-----AEKSESGLVNGYSIYNTGEOSNLDLGRNKAD-GEVIEV 456
 Db 389 RIELSQMGFGQYVNSKSVHEKCYSDGLKLVNKNKTITAGVNSIDESQNNKNHGYINS 448
 QY 457 KTSSINISWYG-----YEHPSFSLAPNOLSQGINITTLYKRTGTE 499
 Db 449 PTDINVNSQNSOYLDNSVMVNGGQKVCSPLSNNGSNMNAALPNQKINVIYSQNSD 508
 QY 500 OMEPVRA-----OGYVNSIKVNTDPPNNVYVYDNEGKLSIYPSNFVD----- 546
 Db 509 --KPEKHADYRKMGVYSHIPYDLPVENYIGDIDPTKPSLLKFPRAEKYGDSIAY 566
 QY 547 -----INSYESTITVQFNSDPELRTTPAFALS-----TGATADD 583
 Db 567 VSEPLNGANVAKLTSYQ--VLKHEVNTQTKYRIRRYATGGDTASIMPHIIGPSGD 624
 QY 584 VISLGYMAE-----VPGSSNPVYMSKDYLLTSEGDYTLMTYFSINNO-----KD 630
 Db 625 LTNEGHNFSVSSRNKMFYOGNNGKYVNLITDSIELPSCQOTILQ--NTNSODLFLDRI 683
 QY 631 EMKKIGSVSKTPEYTHP-----LFEVGHNDOTSTYTLMAHNRVLPDFTLKNLGLPFN 684
 Db 684 EFTSLPSTPTSTNFEPSLEKIIQVNOLEFSSSOTELAH--TVSDKIIDOVYLVKN 741
 QY 685 G-----ELVVVFROTQSS-----SGSLMA-AQETVHIKGEFVYK 719
 Db 742 ALSDQVGEKKAALKRLVYNQAKOLSKARNVLYGSEFKGEMALSRATVYVANHLEKGD 801
 QY 720 PVVEGP---IPDSYRATLHAFYVNGOOLYKGRNTYV--IYNGTAVALIES--SEI 772
 Db 802 HLLPPTLYPSAYQKI-----DESKLKSSTRYTVSGFIIQSEHLEVVYSRYSKEY 853
 QY 773 R-VFPNPARDVVELSAP-----CIPOETSLILD--LSGKIYKNSLSAGHRDVS--RL 823
 Db 854 HDMLDIPYEALPISDESPPNCKPAACQSSCGSGSDSHFYSIDVGSLOSDVNLGI 913
 QY 824 PNGAYILKVDGYTTKINI 841
 Db 914 EFGRIAKRNGFAKISNL 931

RESULT 20
 US-08-296-791-3
 ; Sequence 3, Application US/08296791

```

Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Treccatlo, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1541 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-3

Query Match      2.8%  Score 124; DB 4; Length 1541;
Best Local Similarity 20.4%; Pred. No. 0.93;
Matches 197; Conservative 106; Mismatches 339; Indels 322; Gaps 53;

QY 57 YVYRQGDAREGITSOE-----GSPAYFYVANGNNE--GYALVAD 96
DB 210 FYKKGDNSLILNNEVGNNLKVGDAYTYGIACTP---YKVNHNENGLIGFGNSKEE 266
QY 97 DRIPT-ILAYSPITGRFD--DSMPDNLRMWLIQYDQIG--LILSG----- 137
DB 267 HSDPKGILSQDPLTNYAVLVGDSGSP-----LFYVDREKGMFLGSDYDFWAGYNNKKSQGE 321
QY 138 -----KAOLNEELLRTGEGVAEYFHALMDNGHFANDENRKNQGYPW--NNKEPILLPNCNNAY 191
DB 322 WNIYSQFTKDYLVNKSAGSLIGSKTD-----YSWSNGKRTSTITGGEKSL 367
QY 192 TGCVAATAAOIMRYHSMPLQEG-----SFDYHAGSLY-----GNWSG--- 229
DB 368 NYDLADGDKRKHGKSVTFEGSGTLLNNNIDOGAGGLEFEGDEYKGTSDNTTWKAGAV 427
QY 230 --TFGEATDWMPCNPDLNLTQSQVDAYATL-----MRDVSASYSMSFYENG 276
DB 428 SVAEKGKTYTW--KVHNPQYDLAKI---GKGTLLIVEGTGDNKGSLLKVGCGTYILKQTING 482
QY 277 SGTYSVVYVVAL--RNNFRYKRSQDLHYRALYTSQEHMDIRGELASGRPYTYAGNNOSI 334
DB 483 SGOHAFASVGYSGSTVLVNDKQVDPNSTYFG-----FRG---GR-LDLNNGNSLTF 531
QY 335 GHAFVCDGASDGTGFHFMNGMGVSNCFYKLLLSPTSLGIGGEGIGFTYIOEITG--- 391
DB 532 DHIRID-----DGAHLVHNNMTNASN-----ITITGE-----SLITDPNT 567
QY 392 IEPAKTPAEAGTDALPILALDIEAEYKSEGLANV--GYSIY-----NTGEE 436
DB 568 ITPYNIDAPDEDDNPYAFRRIRKDGOLY-----LNLNENTYVALRKGASTRSELPRNSGES 622

```

```

QY 437 QSN-LDGLYLNKAD-----GEVIEYKSSINISWYGEHEPSES 476
DB 623 NENMLTGKTSDEAKRNVNHNINNERMNGFNGYEGEGKNNGNLNVTEFGKSEQURFL- 681
QY 477 LAPNLSOGINTITLLYRRTGT--EQMEPYRHAQGYVNSIKVNTDP-----NNVYVYD 530
DB 682 -----LTGGINLNGDITVEKGTFLGSRFPFPHARD--TAGISSTKQDPFAENNEVVED 734
QY 531 -----NNEKLSIVPNSFYADLNS-----YE-HSTITVQ----- 558
DB 735 DWINRNFKATYTNVNTGNASLYSGRNVANITSNTITASNKAQVHIGYGTGTVCYRSDYGTG 794
QY 559 -----FNSDPDEIRTPVAFALSTGATADYSLGMYMAVEPGS-----S 599
DB 795 VICTTKLSDKALNSNPINLKNVNLTSANFVLKAKMLFGTISR--GNSQVRLTENS 852
QY 600 NPPVYMSKVY--LTLSEGDYTLMYRFSINNOKDEMKGISVSKPTPEX--THPLFEVGNH 656
DB 853 HWHLTGNSDVHOLDLNGHIL-----NSADNSNNV-----TKNTILVNSLSGN 897
QY 657 QSTYTLDAHNRVLPDFTLKNLGLPENGELVYVFRQTOSSSG--SLMAAQETVHIKQGE 714
DB 898 GSEYTLFDLSNKQ-----GDKTVV--TKSATGNFTLQVADTGPENHNE 939
QY 715 TEVYK-----PYVEGPIDPGSYRATLHAFVNGQOOLYK--GGRNTYKVIYNGT 761
DB 940 LTLFDASKRQDHLNVLGNTVDGAKYKILRN--VNGRYDLINPEVEKRNQTVDTNTIT 998
QY 762 AVEAIE-----SSEEL-RVFPNPARDYVEISAPCIPOETSILFDLS--GRIYAKN 809
DB 999 TPNNIQADVPSVSNNEETIARVDEAP---VPPAPAPRSETTEETVAENSKSEKSTVEKN 1054
QY 810 SLISA 813
DB 1055 EODA 1058

```

Search completed: May 29, 2003, 13:27:45
Job time : 23 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	491	11.1	398	9	US-10-002-78A-24	Sequence 24, Appl
2	393	8.9	468	9	US-10-002-78A-27	Sequence 27, Appl
3	386	8.7	248	9	US-10-002-78A-25	Sequence 25, Appl
4	154	3.5	1355	10	US-09-741-6659-304	Sequence 304, Appl
5	148.5	3.3	1352	9	US-09-784-55AB-2	Sequence 2, Appl
6	140	3.2	934	8	US-08-837-455-19	Sequence 19, Appl
7	138.5	3.1	1391	9	US-10-080-505-11	Sequence 11, Appl
8	138.5	3.1	1391	9	US-10-080-505-15	Sequence 15, Appl
9	138	3.1	1398	12	US-10-090-624-6	Sequence 6, Appl
10	138	3.1	1411	9	US-10-080-505-17	Sequence 17, Appl
11	136.5	3.1	3354	9	US-10-160-758-11	Sequence 11, Appl
12	136.5	3.1	3354	9	US-10-160-758-12	Sequence 12, Appl
13	131	3.0	2993	9	US-09-738-626-6539	Sequence 639, Ap
14	125	2.8	878	10	US-09-912-020-364	Sequence 364, Ap
15	124	2.8	1350	9	US-09-784-55AB-4	Sequence 4, Appl
16	124	2.8	1541	9	US-09-839-996-3	Sequence 3, Appl
17	124	2.8	1541	9	US-10-080-505-3	Sequence 3, Appl
18	122.5	2.8	833	10	US-09-844-281-1	Sequence 1, Appl
19	122.5	2.8	867	9	US-09-839-894-6	Sequence 6, Appl

20	122	2.7	2343	9	US-09-769-787-2	Sequence 2, Appl1
21	121	2.7	2244	10	US-09-815-742-1	Sequence 12713, <i>A</i>
22	120.5	2.7	1599	9	US-10-092-880-9	Sequence 9, Appl1
23	120	2.7	956	9	US-10-121-032-63	Sequence 63, Appl1
24	120	2.7	956	9	US-10-093-037-63	Sequence 63, Appl1
25	120	2.7	1198	9	US-09-975-719-405	Sequence 405, Appl
26	120	2.7	1124	9	US-10-131-591A-75	Sequence 25, Appl
27	120	2.7	5795	10	US-09-815-242-12610	Sequence 12610, <i>A</i>
28	119.5	2.7	1436	9	US-10-080-505-13	Sequence 13, Appl
29	119	2.7	1395	9	US-10-080-505-7	Sequence 7, Appl1
30	117	2.6	1228	10	US-09-137-531-15	Sequence 15, Appl
31	117	2.6	1222	9	US-09-117-447-2	Sequence 2, Appl1
32	117	2.6	1252	10	US-09-137-531-9	Sequence 9, Appl1
33	116	2.6	683	9	US-09-738-626-6685	Sequence 6685, Ap
34	116	2.6	1394	9	US-09-839-996-2	Sequence 2, Appl1
35	116	2.6	1394	9	US-10-080-505-2	Sequence 2, Appl1
36	116	2.6	2353	10	US-09-797-862-33	Sequence 33, App
37	115.5	2.6	719	9	US-10-234-266-2	Sequence 2, Appl1
38	115.5	2.6	4303	9	US-09-904-968A-2	Sequence 2, Appl1
39	115	2.6	1997	9	US-09-909-567B-54	Sequence 54, Appl
40	114.5	2.6	1098	10	US-09-797-862-32	Sequence 32, Appl
41	114	2.6	852	9	US-10-097-340-57	Sequence 57, Appl
42	114	2.6	1042	9	US-10-131-591A-78	Sequence 78, Appl
43	114	2.6	1065	9	US-10-097-340-59	Sequence 59, Appl
44	114	2.6	1069	9	US-10-097-340-55	Sequence 55, Appl
45	114	2.6	1090	9	US-10-097-340-61	Sequence 61, Appl

ALIGNMENTS

```

RESULT 1
US-10-002-784A-24
Sequence 24, Application US/10002784A
Publication NO. US20030036644v1
GENERAL INFORMATION:
    /33
APPLICANT: Ulrich, Robert G.
TITLE OF INVENTION: Bacterial Superantigen Vaccines
FILE REFERENCE: 003/233/SAP
CURRENT APPLICATION NUMBER: US/10/002,784A
CURRENT FILING DATE: 2001-11-26
PRIOR APPLICATION NUMBER: 08/882,431; 09/144,776
PRIOR FILING DATE: 97-06-25; 98-09-01
NUMBER OF SEQ ID NOS: 40
SOFTWARE: Apple Macintosh Microsoft Word 6.0
SEQ ID NO 24
LENGTH: 398
TYPE: prt
ORGANISM: Artificial sequence
FEATURES:
OTHER INFORMATION: mutant streptococal pyrogenic exotoxin B prosegment
US-10-002-784A-24

Query Match      11.1%; Score 491; DB 9; Length 398;
Best Local Similarity   30.6%; Pred. No.-2.2e+29;
Matches     129; Conservative       72; Mismatches   156; Indels        64; Gaps         13


QY          6 LLAIVMLFGIAMQC-----HSAPVTKERALSRLARLROVSLRMGQTAVSDKISIDYYV 59
           |||::|||::|||||::|||::|||::|||::|||::|||::|||::|||::|||::
Db          10 LLSTLLATLGFLPANDVFADONFARNREKAEDSATFFIOKSAAIKGAARSABD-IKLDRVN 68

QY          60 RQGDAERGITSQEESGPAYFYVANNGNEGVALVAADDRIPILAYSPIGRDMSPDN 119
           |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          69 LGGL-----SGSNRTGYNISIG---GFVIYSGDKRSPILDISTGSFDANG-KEN 116

QY          120 LRWLKIIT-----DOEIIILSGRAQLNEELIRTEGVPAEVHALMDGHFPAMDPRNQ 173
           ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          117 IASPMSEIVEQIKEKKRKLDITYAGTAIEKPV-----VKSLDD-----SKGIHYNO 162

QY          174 GFPPNNKERPLPN-----GNHRYTCGVAITAQAQMIRYHSMWLOEGSFDY----- 218
           |:-|-|-|::-|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          163 GNPNALLPTVEKYKPGQRPSVGQHAAAGCAVATAAQIMKHANNPKNLKDVTYTTLSSNN 222
```

OY 219 ----HAGSLVGMGSGTGEYMDWIN-MPGNDPLDNLQSOVDATILMRDVSASVSFY 273
 DB 223 PYFHPKRL---FAAISTROYNNMNLPTYSGRSNOVKM--AISELMADVGISVDYDYG 277
 OY 274 ENGSGTYSVYVVGALRNFRYKRSIQLVHVALYTSQEMHDMINGELASGRPVYAGNNS 333
 DB 278 PSSGSSASSRVOALKEKFGYNSGVHQINQDFFSKODMEQIDKELSONQPVYQGVGVK 337
 OY 334 IGHAFCVGDYASDGTFFHNMKGVSNGFYKLTLLSTSLGIGSGEGFTTYDEITGIE 393
 DB 338 GGAHFVYIDGADGRNFTYHVMWGVSDGFFRLDALNPSALGTGGAGGFGNGYOSAVYGIK 397
 OY 394 P 394
 DB 398 P 398

RESULT 2

US-10-002-784A-27
 Sequence 27, Application US/10002784A
 Publication No. US20030036644A1
 GENERAL INFORMATION:
 /33
 APPLICANT: Ulrich, Robert G.
 TITLE OF INVENTION: Bacterial Superantigen Vaccines
 FILE REFERENCE: 003/233/SAP
 CURRENT APPLICATION NUMBER: US/10/002,784A
 CURRENT FILING DATE: 2001-11-26
 PRIOR APPLICATION NUMBER: 08/882,431; 09/144,776
 PRIOR FILING DATE: 97-06-25; 98-09-01
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: Apple Macintosh Microsoft Word 6.0
 SEQ ID NO 27
 LENGTH: 468
 TYPE: PRT
 ORGANISM: Artificial sequence
 FEATURE:
 OTHER INFORMATION: mutant Spea-mutant Speb fusion
 US-10-002-784A-27

Query Match 8.9%; Score 393; DB 9; Length 468;
 Best Local Similarity 35.0%; Pred. No. 8.5e-22;
 Matches 106; Conservative 43; Mismatches 114; Indels 40; Gaps 13;

OY 111 FEMDSMPD-NLKMMLQIY-DOEIGLISGAQLNEIITGCVPAEVHALLMDNGHFRANDP 168
 DB 184 FDFPEPEPTQSKYLMYKDN---TLDSTQI--EYVLTTRQPV-VKSLD-----SKG 232
 169 MRMNOGYPMNKEPLLPN-----GNHAYTGCVATAAQAQIMRYHSPLOGEGSFYD- 218
 DB 223 IHHNGNPNLTPYIEKKYKPEQSFVGOHAATGCVATATQAIMYHVPKNGKLDYTYT 292
 OY 219 -----HAGSLVGMNSGTGEYMDWIN-MPGNDPLDNLQSOVDATILMRDVSASV 268
 DB 223 LSSNPNFYFHPKRL---FAAISTROYNNMNLPTYSGRSNOVKM--AISELMADVGISV 347
 OY 269 SMSFENSGTYSVYVVGALRNFRYKRSIQLVHVALYTSQEMHDMINGELASGRPVYA 328
 DB 348 DMDGPPSSGSSASRVOALKEKFGYNSGVHQINQDFFSKODMEQIDKELSONQPVYQ 406
 OY 329 GNNOSIGHAFVCDGYASDGTFFHNMKGVSNGFYKLTLLSTSLGIGSGEGFTTYDEIT 388
 DB 407 GKGVG-GHAFFVYIDGADGRNFTYHVMWGVSDGFFRLDALNPSALGTGGAGGFGNGYOSA 465
 OY 389 ITG 391
 DB 466 VVG 468

RESULT 3
 US-10-002-784A-25
 Sequence 25, Application US/10002784A

Publication No. US20030036644A1
 GENERAL INFORMATION:
 /33
 APPLICANT: Ulrich, Robert G.
 TITLE OF INVENTION: Bacterial Superantigen Vaccines
 FILE REFERENCE: 003/233/SAP
 CURRENT APPLICATION NUMBER: US/10/002,784A
 CURRENT FILING DATE: 2001-11-26
 PRIOR APPLICATION NUMBER: 08/882,431; 09/144,776
 PRIOR FILING DATE: 97-06-25; 98-09-01
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: Apple Macintosh Microsoft Word 6.0
 SEQ ID NO 25
 LENGTH: 248
 TYPE: PRT
 ORGANISM: Artificial sequence
 FEATURE:
 OTHER INFORMATION: mutant streptococcal pyrogenic exotoxin-B
 US-10-002-784A-25

Query Match 8.7%; Score 386; DB 9; Length 248;
 Best Local Similarity 36.2%; Pred. No. 1.1e-21;
 Matches 93; Conservative 39; Mismatches 93; Indels 32; Gaps 8;

OY 155 VALMDNGHFRANDPMRMNOGYPMNKEPLLPN-----GNHAYTGCVATAAQAQIMRY 205
 DB 4 VVSLD-----SKGHHYNGNPNLTPYIEKKYKPEQSFVGOHAATGCVATATQAIMKY 58
 OY 206 HSMPLQEGESFSDY-----HAGSLVGMNSGTGEYMDWIN-MPGNDPLDNLQSOV 254
 DB 59 HNYPMKGLKDYTYLLSSNPNFYFHPKRL---FAAISTROYNNMNLPTYSGRSNOVKM- 114
 OY 255 DAYATLMDVSAVSMSFENGSGTYSVYVVGALRNFRYKRSIQLVHVALYTSQEMHDM 314
 DB 115 -AISELMADVGISVMDGPPSSGSSASRVOALKEKFGYNSGVHQINRSD-SDMEQAQ 172
 OY 315 INGELASGRPVYVYAGNNSIGHAFVCDGYASDGTFFHNMKGVSNGFYKLTLLSTSLG 374
 DB 173 IDKELSONQPVYQGVGVK-GHAFFVYIDGADGRNFTYHVMWGVSDGFFRLDALNPSALG 231
 OY 375 IGEGSIGFTTYOEITG 391
 DB 232 TGGAGGFGNGYOSAVYG 248

RESULT 4

US-09-741-669-304
 Sequence 304, Application US/09741669
 Patent No. US20020022718A1
 GENERAL INFORMATION:
 APPLICANT: Forsyth, R. Allyn
 APPLICANT: Ohlson, Karl L.
 APPLICANT: Zyskind, Judith W.
 TITLE OF INVENTION: Genes identified as required for
 FILE REFERENCE: ELIFRA.009A
 CURRENT APPLICATION NUMBER: US/09/741,669
 CURRENT FILING DATE: 2000-12-19
 PRIOR APPLICATION NUMBER: US 60/173005
 PRIOR FILING DATE: 1999-12-23
 NUMBER OF SEQ ID NOS: 481
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 304
 LENGTH: 1325
 TYPE: PRT
 ORGANISM: Escherichia coli
 US-09-741-669-304

Query Match 3.5%; Score 154; DB 10; Length 1325;
 Best Local Similarity 21.7%; Pred. No. 0.0073;
 Matches 157; Conservative 86; Mismatches 260; Indels 220; Gaps 39;

OY 161 NGHFRANDPMRMNOGYPMNKEPLLPN-----PNCNHAYTGCVATAAQAQIMRYHSMPLQEGESFD 217

Db 295 NGFNGNCTVINSNGLINLNKESLVGVQDSHGCVNTDK----- 334
QY 218 YHAGSLVGNWS--CTGCE--MYDMINPBNPDLNLTOSQVADATILMDVSASVMSFY 273
Db 335 -----GHNFLGT--GEAFRIYIGDAGDELNVSSGKYD-----SGITIGMK 377
QY 274 ENGSGTYSV---YVVALRNNFRYKRSLOLVH--RALYTSQEWMDIRGELASGRVYV 327
Db 378 ETGTGNTVYDKMSVITNLGTNLGYDGHGEMNINSGSLVSSSGTGYETGYGNVST 437
QY 328 AGNNOSIG--HAFVCDGYASDGTTFHFMWG-----MGVSNCFYKLTLSPTS--- 372
Db 438 TGMGMVEYKNVYTTIGVAGVGNLNISDCKFVSQNTIFELDKASGIGTLMMDATSSFD 497
QY 373 -----LGIGEGCI-----GFTIYOE---ITIGIEPAKPPAAGTDALPILAKIDEAYKS 420
Db 498 VGINVGNFSGGIYVNSNGATLNSTGYGFIGGNASGKIYVNSTSLMNLSTSTNAOL- 555
421 ESGLNVGYSIYNTGEOSNIDLGYRLNKADGEVLEVTSST-----NISWYGEHP 472
556 ---LQVY--VLGTGE-----LNTTGGIYKADTQIALNDKSGDYRVDGQNSL 600
QY 473 ESFSLAPNOLSOGINTITLLYRRT---GTEQW---EPV-----RHAOGGYV 512
Db 601 ETENMYVG--TSGTGTLTLNNGTLNVEGGEVYLGVEPAVGLNIGAHGEAADAAGFI 658
QY 513 -NSIKV-----NTDPN---NVVTVYDNNEKLSIVPSPFADLNSYHST 554
Db 659 TNAKTFEGLGEGVEFENHTNSDAGYQVDMLTITGDDKDKVI-----HDA 704
QY 555 ITVQFNSDSPDEIRTPAFALSTGA--TADDVISLGWMAEV---PGSSSNYPVWMSKD 608
Db 705 GHVFNAGNTYSKTLVNDGLLTIASHTADGVTGMG--SEVITIANG-----TLD 753
QY 609 VL--TLSEGDYTLMTRES---INNQDEMKKIGSVKPTTEYTHPLFEVGHNOTSTYT 662
Db 754 ILASTNSAGDYTLTNALKGDLKRVOLSSSDKMGFGFHAATGEFA---GVAQLKDSFT 809
QY 663 LDMAHNRVLPDFLTKNLGLPFGELVYVFPROTOSSSSLMAAOETHYIKOGFTYV--- 718
Db 810 LEDNTRALTHAMQOS-----DSBNTTSVAVGEGSISGL-----AMNGGTIIFPTDI 856
QY 719 --KPVVEGPIDPSYRATLHAFVNGQOOLYIKGRNTYVIRVNGTAVEALESSEIRVFP 776
Db 857 PAFTLAEGYI-----SVDTIIVVAGDYTWKG--RANQ---VNTGTGVLD-----VP 898
777 NPARD 781
899 KPMND 903

RESULT 5
US-09-784-554B-2
; Sequence 2, Application US/09784554B
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1352
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-2

Query Match 3.3%; Score 148.5; DB 9; Length 1352;

Best Local Similarity 18.6%; Pred. No. 0.02;
Matches 176; Conservative 127; Mismatches 334; Indels 307; Gaps 47;

QY 112 DMDSPNLNRMLDIYQIEIGL-----ILSGAQLNEELLRECV 151
Db 282 DMSVKNYEMFEDYLDQKRLSSQVEGKRLLDVDFVHWYPEAMGGGIRITNEVGNDETK 341
QY 152 PAETHALMDNGHFANDMPRNNQY-----PMNK--EPLLP----- 185
Db 342 KARQA-----PRLMDPTYKEDSMIAQNSEFLPLIRLKQSVDKYTPETKLAMT 392
QY 186 -----NGNHAATGCVATAAA-----OIMRHSWPLD--GEGSPDYHAGSLVGNMG--T 230
Db 393 EYSYGENDSLSSGIAHMDVIGLIGKNDYVYANMYKLDGVNMYSAAYKLYRNDGNST 452
QY 231 FG-----EMATDNMGPNPDLNLTOSQVADATILMDVSASVMSFYENGSGTY--SVY 283
Db 453 FGDTSVSAQTSDIYNSVHASVYNASDKELH--LVYMKMSDASAFDQFDLSGAKTYISGK 511
QY 284 VVALRNNFRYKRSLOLVH--RALYTSQEWMDIRGELASGRVYVYAGNNOSI---GH 336
Db 512 VMGFDKNSQIKKAPITQISGRNFTYV-----PLTAIYVLTGNDTSPVSGPE 563
QY 337 AFVCDGYASDGTTFHFMW-----GWC-----GVSNG-- 361
Db 564 SFRLKAPAGDOKVHLSPDASGVYGVSVQATDENGFYAVASNLFTSYDTNVTGTS 623
QY 362 -FYKLTILSPSLGIGEGGFTIYOEITIGIEPAKPPAAGTDALPILAKD--IARVY 418
Db 624 YYYKVT--ARTNKSSSSNLKAVKMPVNG--PARAEABEGT-----LKGTVSSG 672
QY 419 KSEGLNVGYSIYNTGEOS-----NLDLGR-----LN--KDGVEIE 455
Db 673 TGISAGAYVNFHNPGLSLYHTIOAPTAGLNLITIGRSPDOKRTNPSLNGKAFGLL 732
QY 456 VKTSINISWYGEHBESESLAPNOLSOGINTITLLYRRTGTBOWEPVRAOQGYNSI 515
Db 733 KRTADF-----KETSGKVLINAGANTISF--ETGM--GWYDI-----DYVRL 772
QY 516 KVNTDNNVYVYDNNEKLSIVPNSFVAD-----LNSYESTITVQFNSDSPDEIR 568
Db 773 PAADRPAAVATKLTN-----PNATVEKALMNTLVDDYGNMLSGOEIETIDWLQ 824
QY 569 TPVAFALSTGATADVVISLGWMAE--VPGSSSNYPVWMSKD--VLTLSBGDYTLWY--- 621
Db 825 ANV--GKRPALADLIDYSPRAHGLSTFEAKALAMKOGGIVTFA-----WIMNA 876
QY 622 -RESINNO--KDEWKKIGSVKPTTEYTHPLFEVGHNOTSTYTLDMAHNRVLPDFLTKNL 679
Db 877 PKGLIDTQGEKEMWGFYADSTTFDIEYA-----MNHBESEDYKLLRIDIVIAQ--QLK 930
QY 680 GLPFGELVYVFPROTOSSSGS--LMAAOETHYIKO----- 712
Db 931 Q--DARVPVLFRLHBAEGKFWMGAKGPEPVKLYILMHDRLTNVHKLNNLIWVNSV 987
QY 713 -----GEPFY-----YKPVV-----EGPIPD-----G 729
Db 988 APDMYRDEYVDILSPSYQAGDYSPQISYEDLVALGDKRKLYVASENKPPIPDPLMK 1047
QY 730 SYRATLHAFVNGQOOLYIKGRN--YVAVINGTAVEALESSEIRVFPNARDYVETIS 786
Db 1048 AYQAHMGMFAMWYDGFVRDQKNSLEHLKRYVNHNPVYITLDE-----LPTNLTGYITE 1101
QY 787 APCIPQETSTILPDLSGKIYKNSLSGHGMVSR--LRPGAY 828
Db 1102 QPSVPGSFTLMAAGETAKVSLSWTASANASYEVKRSRTTEGAP 1145

RESULT 6
US-08-837-459-19
; Sequence 19, Application US/08837459
; Patent No. US20020006407A1
; GENERAL INFORMATION:

```

APPLICANT: McKee, Marian L.
APPLICANT: O'Brien, Alison D.
APPLICANT: Machtel, Marian R.
TITLE OF INVENTION: Histidine-Tagged Intimin And Methods Of
TITLE OF INVENTION: Using Intimin To Stimulate An Immune Response And As An
TITLE OF INVENTION: Antigen Carrier With Targeting Capability.
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: Flanagan, Henderson, Farbow, Garrett &
ADDRESS: Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/837,459
FILING DATE: 18-APR-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Foidis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04995.0023-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 934 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-837-459-19

```

Query Match 3.2%; Score 140; DB 8; Length 934;
Best Local Similarity 17.4%; Pred. No. 0.05;

Matches 167; Conservative 128; Mismatches 331; Indels 336; Gaps 45;

```

74 GSPAYFYVANR--GNNEGVALAADRITITLAYSIGRF-----DMSMPD-N 119
26 GIGLFFVYVONSNFANGENTFKLGSOSKLTHTSYO--NLFYTLKGTGADLSKSDIN 83
120 L-RMW--LQIYDOEIGLILSGKAQLEIEILRTGVPVAVHLMONGHFAANDPMKNGCY 175
84 LSTIMELNKLHLYSSESEMMKAPGO--QIILPLKLLPEYSAL-----124
176 PMNNKEPLLPNG-----NNAVTCGVATAAQIMRYHSMPLQG 212
125 PLLGSAPLVAAGVAGHTNKLTKMSPDYTKSNMTDKALNVAQAQASISGLSRLNG 184
213 EGSFYHAGSLVGNNSGTFGEYMDIINMPGNPDL-----DNLTQOVAVATLMDVSAS 267
185 DYAKDTALG-INGNASS--QIQANLQHYGTAEVNLGSDNFDGSLD-----229
268 VSMSEYEN-----GSGT-----YSYVY-----284
230 FLPLPYDEKMLAFQGVAGARYIDSRFTANLCAQGFLLPAMMLGVNFTDDFGSDNRL 289
285 -VGAIRNNFRYKRSQQLVRLALYTSQEWHDIMRGELASGRVY-----YAGNNSIG 335
290 GIGGEYMDYFKSS-----VNGYFRMRWHEHYHKKDYDERPANGFDIRFNGYLPSPALG 345
336 HAFVCDGASDGTFRHFMNGMGVNGFYKLTLLSPTSIGIGEGITFYOETITTEPA 395
346 AKLIEQYGGDNVALFNSD-----KL-----QSNPGAATVAVNTTPIPLVYMG1---389
396 KTRPAEGDALPILALADIEAEVYSESGLANVGYIYNTGEBOS---NLDLGYRLNKADGE 452

```

```

DB 390 -----D-R-----HGTGENDLISMOFRIOFDSWSQ 417
OY 453 VIEWKTSINISWYGEHEBESFSLAPNLSGINTITLLYR-----TGTEQ 500
DB 418 QIEPQ-----YVNELRILSGSRDYDVOHNNHIIIEYKKODILSINPHDINGTEH 467
OY 501 -----W-EPVRAOGGVYVSVIVNTTDPNNVYVTDNNEGKLSYV 540
DB 468 STQKIDLIYKSKYGLDIRYWDSDALSOGGQIOHSGSOSAODYQATLPAYVGGG-----522
OY 541 NSFVADLNSTYHS-----TITVOFNSDSPDEIRTPVAFALSTGATADVISLGYW 591
DB 523 NIYKVTARAYDRNGNSNNVQLTITVLSNGOVVDQGVYDTPADTSAKADADITITTA 582
OY 592 AEVPGG--SSNPVYVSKVULISBEDYILMTFRFSINNOCKDKKIGSVKTPPEYVHP 649
DB 583 TYKKNVGAQANVPVSN-----IVSGTATL-----GANSKATDANGKATVTLKSTP-----629
OY 650 LEEVGHNOTSTYTLMAHNRVLPDFTLKNLGLPFNGELVYVPROTSSGSLMAAQETVH 709
DB 630 -----GOVVVSAKTAEMSS-----ALNASAVIIPFDQKASITET-KADKTTA 670
OY 710 IKQGE--TFYKKPVVEG-PIPDGSTR-ATLHAFVNGOQQLYL-----K 748
DB 671 VANGKDAIKTYVKVKNNGOPVNNQSVTFSTNGFMNGKSGQATGNDGRATITLTSSA 730
OY 749 GRNNTYKIVNGTAVEALESS--EEIRV-----FPAPADYVEISAPC--789
DB 731 GRATVSATVSDGAEVKATEVFPEDLKDKNVDIIGNNVRELPIWILQYGFLLKASGG 790
OY 790 -----IQETSIILFDLSGKIVAKNSLSAGHRNDVSRPLNG-----AVILKYDGYTTK 838
DB 791 DGTYSWYSEMTSIATVDASGKVTIN-----GKGSV-VIKATSGDKQVYTYTIKAPSWIK 844
OY 839 IN 840
DB 845 VD 846

```

RESULT 7

US-10-080-505-11
Sequence 11, Application US/10080505
Publication No. US20030073166A1

GENERAL INFORMATION:
APPLICANT: St. Gene, Joseph W.

TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
FILE REFERENCE: A-59941-1/REF/DOE/DHR

CURRENT APPLICATION NUMBER: US/10/080, 505
PRIOR FILING DATE: 2002-02-22

PRIOR APPLICATION NUMBER: US 08/296,791
PRIOR FILING DATE: 1994-10-25

PRIOR APPLICATION NUMBER: US 09/839,996
PRIOR FILING DATE: 2001-04-20

NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1

SEQ ID NO 11
LENGTH: 1391

TYPE: PRP
ORGANISM: Haemophilus influenzae

US-10-080-505-11

Query Match 3.1%; Score 138.5; DB 9; Length 1391;
Best Local Similarity 20.4%; Pred. No. 0.12;

Matches 165; Conservative 89; Mismatches 275; Indels 281; Gaps 44;

```

OY 67 GITSQEGSPAIR-----YVANGNNEGVALAADD-----RIPTILAYSPIGRPD 112
DB 18 GIVSQAMAGHTYRGIDYQYRDFAEKKGKFTYGAQOIIDYNNKGEMIGTMKGVMP--D 75
OY 113 MDSM-----PDNLRMLQI---YDOETGLIL 135
DB 76 LSSMAYGYSITLISEOHLISVAHNVGYDVYDFGMEGENDDHFRFYKVKRYNRYK-----130

```

Query 136 SGAOLNE-----EILRTGVPAAEVAHALDNGHPAND-----PMRNOCYPM-----NNK 180
Db 131 SGRQNDYDHPLEKFTETAPLEIEMSYDGNHKNFNQYPLRVRGSHQWKKDDNNK 190
QY 181 EPLLPNGNHAATGCVATAAQMIRYHSMPLQEGSFYDAGSLVGNMSTFGGEMDWINM 240
Db 191 -----TIGDLAYGG-----SMLIGMTFEDGPGAGN----- 215
QY 241 PGAPDNLQSOVDAYATLMDVSAVSMSFENGSGTYSYVVGALRNFRYKRSLOL 300
Db 216 -GTLELNGRVQNP-NKYGPL-----PTAGSFDSGSPM-----FYDKYV-- 253
QY 301 HVALYTSQEMHDMIRGELASGRPVYAGNNSIGHAFVCDGYAS--DGEHFPMWGCVS 359
Db 254 -----KKW--LLNGVLREGNPPYAAVNSYQITRKDYFOGILNODITANF--WD--T 298
QY 360 NGFYKLLTSPSLGIGEGIGFTTYOET-ITGIEPAKTPAEAGTALPILAK--DIEA 416
Db 299 NAEYRF-----NIGSDHNGRVATIKSTLPKKAIOPERIYGLVDSQLHDARDKNGDESP 352
QY 417 EYKSES-----GLNVCYSTYNTGEEQSNLDGYRLNKADGEV-----LEVTSINISWYG 467
Db 353 SYKGPWPSPALHGHKSIYFGDGTGTLTIENNINOGAGGLYEGNFFVVGNGNONNITWOG 412
QY 468 YG-----EHPESFSLAP-----NOLSGINTITLLYRRGTE 499
Db 413 AGVSVGESEVEMOVHNPEDGRSLKIGLTLVNGKKNIGSLSVGNGLVVL--DQADE 470
QY 500 QWEPVRAAGGYVN--SIKVT--TDPNNVVY--TVDNNEGLSTI--VPN--S 542
Db 471 SGQKQAFKEVGIYSGRATVOLNSADQVDPNNIYFGRGRLDNGHSLFERIQNTDEGA 530
QY 543 FVADLNSYEHSTTV-----QFNSDSPDEITPVAFALSTATADVLSLWV--MAEV 594
Db 531 MIYNHNASQJANTITIGMATINDSKOL-----TNKKDIAFNCGFEGODAKT 578
QY 595 PGG-SSNYPYVWMSKDYLTLSGDDYTLMYRESINNOCKEMKIGSVS--KTPTETHP 649
Db 579 NGRLVNHYQPVNAENHLLSG-----TNLNGNITONGTGLVFSGRPPHAYNH- 627
QY 650 LFEVGHNGTSTYTLDMANHRLP-----DFTLKNLGLFPNGELVYVFRQTOSSS 698
Db 628 -----LRRDLSNMEGIFQGEIYWDHDMINRFEKAENFOIKGSAAVSHVSSIE 676
QY 699 GSILMAQETVHIKQGETFYVKKPYVEGPIPD 728
Db 677 GN-W-----TVSNNAATF-----GVVFN 694

DLT 8
US-10-080-505-15
; Sequence 15, Application US/10080505
; Publication No. US2003073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080, 505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/286,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent version 3.1
; SEQ ID NO 15
; LENGTH: 1391
; TYPE: PRP
; ORGANISM: Haemophilus influenzae
US-10-080-505-15
Query Match 3.1%; Score 138.5; DB 9; Length 1391;

Best Local Similarity 20.4%; Pred. No. 0.12;
Matches 165; Conservative 89; Mismatches 275; Indels 281; Gaps 44;
Query 67 GITSQEBGSPAYF-----YVANGNNEGVALVAD-----RIPTLAYSPIGRPD 112
Db 18 GIYSQAMAGHTYTGIDYQYRRDEAENKGFYGAODIDIKKKGEMGTMMKVPMP--D 75
QY 113 MDNM-----PDLNMMWLOI--YDQELGLL 135
Db 76 LSSMVRGYSSTLISEOHLISVAHNGVDYVDFEGMEGEPDQHRFKYVARYNKK----- 130
QY 136 SGAOLNE-----EILRTGVPAAEVAHALDNGHPAND-----PMRNOCYPM-----NNK 180
Db 131 SGRQNDYDHPLEKFTETAPLEIEMSYDGNHKNFNQYPLRVRGSHQWKKDDNNK 190
QY 181 EPLLPNGNHAATGCVATAAQMIRYHSMPLQEGSFYDAGSLVGNMSTFGGEMDWINM 240
Db 191 -----TIGDLAYGG-----SMLIGMTFEDGPGAGN----- 215
QY 241 PGAPDNLQSOVDAYATLMDVSAVSMSFENGSGTYSYVVGALRNFRYKRSLOL 300
Db 216 -GTLELNGRVQNP-NKYGPL-----PTAGSFDSGSPM-----FYDKYV-- 253
QY 301 HVALYTSQEMHDMIRGELASGRPVYAGNNSIGHAFVCDGYAS--DGEHFPMWGCVS 359
Db 254 -----KKW--LLNGVLREGNPPYAAVNSYQITRKDYFOGILNODITANF--WD--T 298
QY 360 NGFYKLLTSPSLGIGEGIGFTTYOET-ITGIEPAKTPAEAGTALPILAK--DIEA 416
Db 299 NAEYRF-----NIGSDHNGRVATIKSTLPKKAIOPERIYGLVDSQLHDARDKNGDESP 352
QY 417 EYKSES-----GLNVCYSTYNTGEEQSNLDGYRLNKADGEV-----LEVTSINISWYG 467
Db 353 SYKGPWPSPALHGHKSIYFGDGTGTLTIENNINOGAGGLYEGNFFVVGNGNONNITWOG 412
QY 468 YG-----EHPESFSLAP-----NOLSGINTITLLYRRGTE 499
Db 413 AGVSVGESEVEMOVHNPEDGRSLKIGLTLVNGKKNIGSLSVGNGLVVL--DQADE 470
QY 500 QWEPVRAAGGYVN--SIKVT--TDPNNVVY--TVDNNEGLSTI--VPN--S 542
Db 471 SGQKQAFKEVGIYSGRATVOLNSADQVDPNNIYFGRGRLDNGHSLFERIQNTDEGA 530
QY 543 FVADLNSYEHSTTV-----QFNSDSPDEITPVAFALSTATADVLSLWV--MAEV 594
Db 531 MIYNHNASQJANTITIGMATINDSKOL-----TNKKDIAFNCGFEGODAKT 578
QY 595 PGG-SSNYPYVWMSKDYLTLSGDDYTLMYRESINNOCKEMKIGSVS--KTPTETHP 649
Db 579 NGRLVNHYQPVNAENHLLSG-----TNLNGNITONGTGLVFSGRPPHAYNH- 627
QY 650 LFEVGHNGTSTYTLDMANHRLP-----DFTLKNLGLFPNGELVYVFRQTOSSS 698
Db 628 -----LRRDLSNMEGIFQGEIYWDHDMINRFEKAENFOIKGSAAVSHVSSIE 676
QY 699 GSILMAQETVHIKQGETFYVKKPYVEGPIPD 728
Db 677 GN-W-----TVSNNAATF-----GVVFN 694

RESULT 9
US-10-090-624-6
; Sequence 6, Application US/10090624
; Patent No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshi
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090, 624

CURRENT FILING DATE: 2002-03-06
 PRIOR APPLICATION NUMBER: 09/445,472
 PRIOR FILING DATE: 1999-12-06
 PRIOR APPLICATION NUMBER: 151969/1997
 PRIOR FILING DATE: 1997-06-10
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 6
 LENGTH: 1398
 TYPE: prt
 ORGANISM: Pyrococcus furiosus
 US-10-090-624-6

Query Match 3.1%; Score 138; DB 12; Length 1398;
 Best Local Similarity 19.5%; Pred. No. 0.13;

Matches 157; Conservative 91; Mismatches 262; Indels 296; Gaps 41;

185 PNGNHAYTCVATAAAQIMRHSWPLOGSGSPDYHAGSLVGNWMSGTFGEYMDINMPNP 244
 351 PGEYAVFEG-----WDHGHGT--HVAQYAGYDSN--NDAMPMLSYSGE 392
 245 -----DIDNLTQ-----SOYDAVATLMRDYAS----- 267
 393 WEVFSRLXGMDYNTVTTDVGAVAPCAQIMAIRLSDGRGSMWDIEGMYAATHGADY 452
 268 VSMSP-----YENGSGTYSVY-----VG 286
 453 ISMSLGNAPFLDGDIDPESVANDELKEKGVYVIAAGNEPCINVSQVATKATITYG 512
 287 ALRNNFRKRSIQILHVRALYISQEHMDIRGELASGRPYV-----AGNNOSIGHAFVC 340
 513 A-----AAVPINV-GYVYSQ-----ALGPDYGFYFPATYRIAF-FSS 552
 341 DGYASDGFHFHW-----GMGVSNGFYKLLLSPTSLSIGGEGIGFTYGEITIGEPAT 397
 553 RGPRIIDGELKPNVAVPGRG-----YSSLSLPMWIGG-----ADPMSGTSMAT 593
 398 PAEACTDALPILALDIEAEYKS-----ESGLNVGYSIYNTGEEQSNLDLGYR----- 445
 594 PHSVGVALLISGAKEGIIYVNPDIKKVLESGAWMBEDPYTGQKYTELDGHHGLVNT 653
 446 -----LNKADGEVIEVKTSSINISWYGEHPESTSLAPNQLSGINTITLLYRTGIE- 499
 654 KSMELKAIINGTTLPIVDHMAKSYSDFAEY-----LGVYIRGJLYARNSIPD 701
 500 --QMEPVRAAGGYNSIKVNTTDP-----NNVVYVDNNEGKLSIYNSFVADLSYE 551
 702 IVEWH-IKVGDEIRTEIRTEIYATEPMIKRPFVSGVILENTEFLVY-----KYDEGLE 755
 552 HSTIIVQFNSSDPDEIRTPVAFALSTGATADDVVISLGVMMAEVPGGSSNYPVWASKDYL 611
 756 PGLVYGRILIDDP--TTPV-----IEDEIL-----NTIYIPK---F 787
 612 LSEGYTL-WYRFSLNNQKDEWKIKGSVSKTPTETHPLEFVGNQSTIYLDMAHNV 670
 788 TPENNYTTLWY-----DINGPEVTHHFTVPBGVYLAAMTYWYG 830
 671 L--PD-----FTLKNL-----GLPFNGELVVV--FROGSS-----SGSLW 702
 831 LYRPGMVFYQDLYLPAAVSNRPMGMLWMTGFNAPLYESGFLVRIYVELTTPSW 890
 703 AAQET-----VHIKQETEVYKPEVGE--PIPGSYRATLHAFAVNGOOLYLGKRN 753
 891 YINRYLDTNTEFSIEFMITNIYAPINATLLPIGIGYNASVEY--GDGEFFIKG----- 944
 754 TVKIVNGAVAEIASESEIRVFPNPARDYVELSACIF-QEISTILFPLSGIYVKNKSL 812
 945 -----IEVPESTAEK-----IRGNPSVNSDLDTLYDSKGLVALDGNP 986
 813 AGHRMDYSRLPNGAYILKVDGYTRK 838
 987 TAEEVVEYPRKPGYSIVHGYSVR 1012

RESULT 10
 US-10-080-505-17

; Sequence 17, Application US/10080505
 ; Publication No. US20030073166A1

; GENERAL INFORMATION:

; Applicant: St. Geme, Joseph W.

; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS

; FILE REFERENCE: A-59941-1/REF/DCF/DHR

; CURRENT APPLICATION NUMBER: US/10/080,505

; CURRENT FILING DATE: 2002-02-22

; PRIOR APPLICATION NUMBER: US 08/296,791

; PRIOR FILING DATE: 1994-10-25

; PRIOR APPLICATION NUMBER: US 09/839,996

; NUMBER OF SEQ ID NOS: 58

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 17

; LENGTH: 1411

; TYPE: prt
 ; ORGANISM: Haemophilus influenzae
 US-10-080-505-17

Query Match 3.1%; Score 138; DB 9; Length 1411;
 Best Local Similarity 18.8%; Pred. No. 0.13;

Matches 180; Conservative 124; Mismatches 325; Indels 330; Gaps 48;

1 MKR-----SFLAIWLEGIAMQ-----GHSAPYTERALSLARLARQVSLRGQAVSDK 52
 1 MKRTYERLMLFACISL-GIVSQAMAGHT-----YFGIDYQYRDFENRGRKTYGAK 52
 53 ISIDYVYRQGDAREGITSQEEGSPAFYVYVANGNNEGYALVAADRIPIT--LWASPIG 109
 53 -NIEVYNNKGNLVG--TSMTKAPMLDEFSYR--NGVALVGDQYIVSAVAINCVTAND 106
 110 RFDMDSPDLRLMLOI-----YDQELGLISGKAQLEIILTEGVPAEYH 156
 107 FGAEGQNPQOHFTFYKIVYRNNYKNDQTHPYEKDY-----HNPRLHKEV--TEAPRIDWT 159
 157 ALMDNGHFAN-----DPMKMNQGYR-WNNKEPDLRNGNHAHYGCVATAAAQIMRHSWP 210
 160 SDMNKNKTYDRKYERVRNIGSGWQFWRNDQ--NNGD----- 194
 211 QEGSGFDY-----HAGSLVGNMSGTGEYMDINMPNPDLNLTOSQVDAVATLMRD 263
 195 QVAGAYHYLLAGTINOGAGGWSLSGD--YRQGN-----YGPi--- 234
 264 VSASVMSFYENGSGTYSYVVGALRNNFRKRSIQILHVRALYISQEHMDIRGELASGR 323
 235 ---PIAGSGDSGSEMF-----YDAEKOKWMLINGVLRGN 267
 324 PYVYAGNN--OSIGHAFVCDGYASDGFHFNMGMGVSNGFYKLLLSPTSLSIGGEGSI 380
 268 P--WAGTENTFOLVAKRSFDEILEKOLRTSF--YSPSGNGAYTIT--DKGDS 314
 381 GFTIYO-----EITIGIEPAKTPAEAGTDALPILAKDIEAEYKSES--GLNVGYSIY 431
 315 GIVKQQTGRPSVVRIGLMDOKLPAEG-----KDDVYQGOBNITLPLNNGNLY 364
 422 NTGEQSNLDGRLANKADG--EYIEVKTSSINISWYGEHPESTSLAPNQLSGGIN 487
 365 FGDKNGTIVTLSTINOGAGGLYFEGNFTVSSENNATNOGAGVHGEDS----- 413
 488 TITLLYRRGTGQ-----WEPVRAAGGYVNSIKVNTTDPNNVVYVDNNEG-----K 535
 414 --TYWKVNGVENDRLSKIGKTLHVAKGKNGKISVGDGVILLEQADQKQAOASE 471
 536 LSIYNSFVADLNSYENHSTIYQFNSD-----SPDEIRTPV 571
 472 IGLVSGRGVYQINDOK-----QFNTDKFYGFGRGRDLNGLHSIFKRIQNTDGGATIV 525
 572 AFAISTGATADDVVISLGVMMAEVPGGSSNYPVWASKDYLTLTSEGYTLTMYRESIN--NQ 628

Db 526 NNNATTESTV-----TITGSDTINDNTGDLTKNRDIAFNMGFGD 564
QY 629 KKKMKIG---SVSKPTETHTHLEFVGNQSTYLLD-----MAHNRVLPDF 674
Db 565 KDDTKMTGRLNTYPLNDLNHLLSGGNLKNITQDGTTFVSGRPHPAHNL---N 621
QY 675 TLKNIILPENGELV-----VFROTSSSGSLMAAOFTHIKOGETFYKRV--VEGP 725
Db 622 RLNEJGRP-KGEVYIDDDMINRFX-----AENFOIKGSGTYSRVNVSIEGN 668
QY 726 -IPDGSYRATLHAFFVNGQOOLYLKGRN--YTVKIVNGAVERAIESESEIRVPNPARDY 782
Db 669 WTISNNANATFGVVPNOQNTICRSDMTGLTCTKTVNLTKDKVIDS-----714
QY 783 VESAPCIPOETSIILFDLSGKIYMKNSLSAG-HGRMDVSRLPNGAYILKVDGTYTKIN 840
Db 715 -----IPTT-----QINGSTINLTNNATVNIHG-----LAKLNGNVTLIN 748
Query Match 3.1%; Score 136.5; DB 9; Length 3354;
Best Local Similarity 20.1%; Pred. No. 0.64;
Matches 199; Conservative 126; Mismatches 338; Indels 327; Gaps 52;
21 SAPVTERALSLARLALROVSLRMGO-TAVSDKISIDYVYRQDAERGITSOE---GS 75
344 NAEFNSSEYVAITELAQVGFALPFIQVYDK-----DENGLNSMEFVLYGN 393
QY 76 PAYFYVANGNNEGVALVAADHIPTITLAVSPIGRDMD-----SNPDI---RMLQIY 127
Db 394 NSHHFISPTSYOGKADIRIVAI--LDYETVDARDFDLFANESVDPDHGVAKVITLI 451
QY 128 DOEIGILSGKQNLNEILLTEGVPRAVHALM--DNG-----HEANDPRMNGRP 176
Db 452 NENDNRPISQPLIYNSLENTVGTSVLTATLDADAGTEGVSSTFSDPPRFSL---508
QY 177 MNKKEPLDNGNHAHYTCVATNA---AQIMRYHNPLOGEGSFDYHAGSLVNMWSTFG 232
Db 509 --DKD-----TGLMLLARLDYELIQRTLIIANDDGGEEFTGVRIN-----550
QY 233 EMDYDINMGONPLDNLTOQVDAIYATLARDVSASVSM-----SEFENGSGTYSVY 283
Db 551 -VLD-VN-----DWPPTFOKDAYGALRENEPSYTOVLRLRATDEDSPPNNQITYSIV 601
QY 284 VVGCALNNPFYKRSGLUHVRLATYSOEMHDMINGELASGRPVYVYAGNOSIGHAFVODGY 343
Db 602 SASAFSGSYDI-----SLTEGY-----GVISVSRLPYE-----630
QY 344 ASDGTFHFMWGWGVNGFYKTLILSPISLGIGEGIGFTIYOIITGIEPAKTPA---399

Db 631 -----QISNGILYILTVM---AMDAGNPPLNSTVPVITIEVFDEBNPPTFSKP 674
QY 400 -----EAGTDLPLIALK-DIEAEKSSSGLANVGSITYNTEBOSNLDGRLNK 448
Db 675 AYFVSVENIMAGATVLFNATDLDLRSREYGOBS---IYSL---EESTO-----PRIN 723
QY 449 ADEEVI-----EVTSSINISMY---GYGEHPESPESLAPNOLSGINT--ITLLYRRT 496
Db 724 RSEIITTSILDBETSEVILYIYRAVDGVGHHOKT-----GIATVNTLLDIND 773
QY 497 GTEBOWEPVRAHOGGYNSIKRNTDPN-NVYVTVDNNEK-----LSIYPSFVADLNSY 550
Db 774 NHPTMDAPY-----YINLVEMTPPDSQVTVVAVDDPLGENGTLVYSIQPNKFYSINST 829
QY 551 --EHSITVOFNSDSD-----EIRPVPALSTGATADVLSL-----587
Db 830 TGRIRTHAMLDRENDDPHEALMRRIVSVTCGGRPLKATSSATVFNVLDLNDNDPT 889
QY 588 -----GWMAFVPGSSNYPVWMSKDVLTLESDVTLWYRPSINNOKDEMKGISVS 640
Db 890 FQNLFPVAEVLBEIPAGVSIYOV-----AIDDEG-----GLVSY 927
QY 641 KPTETHTHLEFVGNQSTYLLDMAHNRVLPFTLK---NLGLPENGELVYVFKOTOS 696
Db 928 RMPVGRPMDFLINSSGVVVTTELDERIAEYQLNVAVDAGTP-----TKS 976
QY 697 SSGSLMAAOFTHI---KQETP---YKRPVYEGRP-----DGSYRATLH 736
Db 977 STSTL-----THVLVDNETPFPFAVYVVSSEDPREFRVWMLCTDNDGGLNAELS 1031
QY 737 AFVNG--QOOLYKGRNTYKIVNG-----TA-----VEAIESSEIRVPNPARDYV- 783
Db 1032 YFTTGGVNDGKFSVGRDAVRYVVGIDRTTYAKMILLEAIDNGPRKGRHGTATYFVY 1091
QY 784 -----EISAP-CIPOETSIILFDLS---GKI--YMKNSLSAGHG---R 817
Db 1092 VLDVNDNRPFILOSSYEAQVPEIDIPGHSTIOLKATPADAGEGGRVWYRIILHGNHGNFR 1151
QY 818 MDYS-----RLPNGAYILKVDGY 835
Db 1152 IHVSNGLMGRPPRLDRERNSSHVLVEAY 1181

Query Match 3.1%; Score 136.5; DB 9; Length 3354;
Best Local Similarity 20.1%; Pred. No. 0.64;
Matches 199; Conservative 126; Mismatches 338; Indels 327; Gaps 52;
21 SAPVTERALSLARLALROVSLRMGO-TAVSDKISIDYVYRQDAERGITSOE---GS 75
344 NAEFNSSEYVAITELAQVGFALPFIQVYDK-----DENGLNSMEFVLYGN 393
QY 76 PAYFYVANGNNEGVALVAADHIPTITLAVSPIGRDMD-----SNPDI---RMLQIY 127
Db 394 NSHHFISPTSYOGKADIRIVAI--LDYETVDARDFDLFANESVDPDHGVAKVITLI 451
QY 128 DOEIGILSGKQNLNEILLTEGVPRAVHALM--DNG-----HEANDPRMNGRP 176
Db 452 NENDNRPISQPLIYNSLENTVGTSVLTATLDADAGTEGVSSTFSDPPRFSL---508
QY 177 MNKKEPLDNGNHAHYTCVATNA---AQIMRYHNPLOGEGSFDYHAGSLVNMWSTFG 232
Db 509 --DKD-----TGLMLLARLDYELIQRTLIIANDDGGEEFTGVRIN-----550
QY 233 EMDYDINMGONPLDNLTOQVDAIYATLARDVSASVSM-----SEFENGSGTYSVY 283
Db 551 -VLD-VN-----DWPPTFOKDAYGALRENEPSYTOVLRLRATDEDSPPNNQITYSIV 601
QY 284 VVGCALNNPFYKRSGLUHVRLATYSOEMHDMINGELASGRPVYVYAGNOSIGHAFVODGY 343
Db 602 SASAFSGSYDI-----SLTEGY-----GVISVSRLPYE-----630
QY 344 ASDGTFHFMWGWGVNGFYKTLILSPISLGIGEGIGFTIYOIITGIEPAKTPA---399


```

1  APPLICANT: MIZOGUCHI, HIROSHI
2  APPLICANT: ANDO, SEIJO
3  APPLICANT: HAYASHI, MIKIRO
4  APPLICANT: OCHIAI, KEIJO
5  APPLICANT: YOKOI, HARUHIKO
6  APPLICANT: TATEISHI, NAKOJO
7  APPLICANT: SENOH, AKIHIO
8  APPLICANT: IKEDA, MASATO
9  APPLICANT: OZAKI, AKIO
10 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
11 FILE REFERENCE: 249-125
12 CURRENT APPLICATION NUMBER: US/09/738,626
13 CURRENT FILING DATE: 2000-12-18
14 PRIOR APPLICATION NUMBER: JP 99/377484
15 PRIOR FILING DATE: 1999-12-16
16 PRIOR APPLICATION NUMBER: JP 00/159162
17 PRIOR FILING DATE: 2000-04-07
18 PRIOR APPLICATION NUMBER: JP 00/280988
19 PRIOR FILING DATE: 2000-08-03
20 NUMBER OF SEQ ID NOS: 7059
21 SOFTWARE: PatentIn ver. 3.0
22 SEQ ID NO 6239
23 LENGTH: 2993
24 TYPE: prt
25 ORGANISM: Corynebacterium glutamicum
26 US-09-738-626-6239

```

Query Match	3.0%; Score 131;	DB 9;	Length 2993;
Best Local Similarity	19.5%;	Pred. No. 1.4;	
Matches 150;	Conservative 87;	Mismatches 258;	Indels 274; Gaps 37
QY	136	SGKAQNEELLRTGCVPAEVAHLDNGHFANDPMKRNQGYPMNNKEPLLPGNHA--YTG	193
Db	651	SGRSQLGADIHEIDNSFAKAGRLLD--EVAD-----ETVAQARDEIIAIGTKAKYFG	704
QY	194	CVATRAAALIMYHSW-----PLOGEGSFYHAGSLVGNWSGTCEMTD-----	236
Db	705	DIGS--MTYEOMLNRYLELSGPVDQW-----IDASWAFRQMLEREARLIE	751
QY	237	-----WI-----NMGPAP-----DL	246
Db	752	QDHQGFPSLTVEGVDKLVAAVYHAATDILLPADVAMFLCLRTPGFVNFVPIYDKOV	811
QY	247	-----DNLTQSDVAYATILMDVSAVSMSFYENGSGTYSVYVGALRNFRYKRSIQ	300
Db	812	RRMRSDSLMOQSHDRYADQVALIPGV-----VAVAGITKANEPVADLLDR	858
QY	301	HYRALYTSQEHDM-----INGELASGRVYVYAGNN-----QSIGHAFPCDGYAD--GTF	349
Db	859	FVDDTTRIDHDHRSRDIMGKVLSSPQTEFAGRNIPBSVHSLHA---DMSRSRFEAF	915
QY	350	HFNMG---MGVSNCFYKLTLLSPTSIGIGE--GIGFTYQEIITIGIEPAKPAEAGTD	404
Db	916	HSPTGALVYEDAEHAMLVPLAGSTAFGTTAEKIRPT-----SPDIALPS	962
QY	405	ALPILALKIDAEYKESSEGLNVGYSI---YNGEESQNLDLGYLRNKADGEVITEVKTSS	460
Db	963	AVPLVTCEDAEAMGELTRIAAGCTLTAVNNGTATWETSVADGV-----	1006
QY	461	INISWYCGHPRESFSLAPNOLSGINTITILYKRTGE-----QWEYRIHAQGGYNSI	515
Db	1007	--IADY-----NNVYAGYVPASVVAHPHTADVLVGRAMPVFAA---YKSA	1047
QY	516	KVNTDPNNVYVTVDNNEGKLSIYPSNFSVADLINSYESTITVQPSNDSPEIRTPVAFAL	575
Db	1048	VIGCTDSASVY---EGMLSLV-----HLEHHIV---LKSVDPPD-----GAL	1083
QY	576	STGATADAVIS---LGMVAEYVGGSSNYPVWSKDVLTLSBGDTYLTWRFSINNOK-	629
Db	1084	KVSATADEVVDLGLRLVIVRAEIADEAGNL-----IATLAE-----RFAIRGRKG	1129
QY	630	-----DEMKGIGSVKTPLEIYHPLFEVGHNGTSTYLLDMAHNRV--LPD	673


```

OY 213 EGSEFDYHAGSLVGNMSG---TFGEYMDWIMNPGNPDLDNLTQSOVDAYAT-----259
DB 432 ANNYSAAYKLYKRYDOCKSTFGD-----ISVHAQTSDDLVNNSVHASTYDASKELHLV 485
OY 260 -LMRDVSASVSMSEFYENGSGTY-SVYVYGLARNNFRKXSLQHLV-----RALYT---SOE 310
DB 486 VMKSMDSADAFDLSGETTYSGKVMGFDKNSOIKEAAPITQISGNFFYTVPLPLTA 545
OY 311 WHDM-----IRGELASGRPVYVYAGNNSIGHAFVCDGVSDDCTF-----349
DB 546 YHIVLVLGNDTPVENPESFALRAEADGKSIYLDASSGVVGVORATY-ENGPEAAVAS 604
OY 350 ---HEFMWGVANG---FYKLLPSLPTSLGIGEGIGFTIYOEIITGIEPAKTPAEAGT 403
DB 605 NLVETSYDTNVTNGTSYKKT--AKTKGTASVNLKAVPAPVPG--PDVEAEDGT 660
OY 404 DALPILALDIEAEYSESLANGY--SIYNGEEO-----NLDIGY-----445
DB 661 ---LGTVESSGTFSGTGYVTFNMGDSLMTIQAFTAGLYNLITGYSRPHD 713
OY 446 ---LN-KADGEVIEVKTSSINISWYGEHSEFSLAPNLSOGINTITLLYRTGT 498
DB 714 KRTNFSLNKASGELYLMKTADR-----KTSCKVLLNCAITIGF---ETG-758
OY 499 EOMEVPRHAOGY-VNSIVYNT--DPNNVYTVDNNEGKLSIYVNSFVAD-----L 547
DB 759 --W-----GWYDIDYVLEPADRPRAVTKTLIN-----PNATVEAKALNLYLV 801
OY 548 NSEHSTIVQFNSDPDEIRTPVAFALSTGATADYISLGWMAVYPPGSSYVYVWVK 607
DB 802 DOYGRKMLSGO-----EMPEIDWLQANV-----GKKPAIAL 834
OY 608 DVLTLSEGDYTLMYRESINNOCKKIGSVKPTETHTP--LEFEVG-----HNO 657
DB 835 DLIDVSPRAEHLSTETEKALIMDKOGGI-VTFAMHNPAPKGLIDTGCKEMWGFYAD 893
OY 658 TSTYTLDMAHNRYLPFTLKNGLPENGELVYVFGTOSSSGSLAAOET-----VH 709
DB 894 STTFDELYAMNH-----PESEDKLLIRIDIVIDYAGOLKLODAKVPLEPRHL 941
OY 710 IKOGEFYVVKPYVEGPIPDGSYRATLHAFVNGOOLYLCKRNYTKIYNGTAVEAIES 769
DB 942 EADGKRTWVG--AKGPEYKILYILMH-----DRLTNVHKLANLWVNSVAPDWPYGD 993
OY 770 EDIRV-----FPNAPADYVETAPCIPOETSIILFDLSGIYV 806
DB 994 EYVDILSFDSPQ-AGDY-----SPQAKYEDLVLTGKDKKLY 1030

```

09-839-996-3
 Sequence 3, Application US/09839996
 Publication No. US20030009010A1
 GENERAL INFORMATION:
 APPLICANT: St. Geme III, Joseph W.
 TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/839,996

```

FILING DATE: 20-Apr-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Treacat, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1541 amino acids
TYPE: amino acid
TOPOLOGY: unknown
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-839-996-3

```

Query Match 2.8%; Score 124; DB 9; Length 1541;
 Best Local Similarity 20.4%; Pred. No. 1.8;
 Matches 197; Conservative 106; Mismatches 339; Indels 322; Gaps 53;

```

OY 57 YVYRGDADRGITSGEE-----GSPATFYVANRGNNE-GYALVYAD 96
DB 210 FYTKKGDNTSLLNHNEHGNNLKVGDATYTYGIAGTP--YKVHNEHNGLIGFENSKEE 266
OY 97 DRIFT-ILAYSPIGFEDM--DSMPDNLKMWLOIYDOETG--LILSG-----137
DB 267 HSDPGLIISDPLTYVAVLGDGSP-----LEVYDREKGMWFLDSYDFWAGYNNKKSWE 321
OY 138 ---KAQNEELLREGEVPAEYHALMDNGHFANDMRKNNGIYV--NKEPLPNGNHAY 191
DB 322 WNIYISOFTKVDVNLNDSAGSLIGSKTD-----YSWSNCKTSTITIGEKSL 367
OY 192 TCGVATTAQAQIRYHSMPLQEG-----SPDYHAGSLY-----GNSSG---229
DB 368 NVDLADGDKDPKHGSIYFEGSGITLNNNDGAGLFEEDYVYKSTSDNTTKKGVY 427
OY 230 --TFGEYMDWIMNPGNPDLDNLTQSOVDAYATL-----MRDVSASVSMSEFYENG 276
DB 428 SYAEGKTYTW--KVHNPQDRLAKI---GKTLIVEGTGDNKSLKVGDTYILKQING 482
OY 277 SCTSVYVYVGL--NNNRKRSLOLHVRALYTSQEHMDIRGELASRPYIYAGNNOST 334
DB 483 SGQHAFAVGIYSGRSTVLVNDKQVDPNSIYFG-----FRG---GR-LDLNNGNSLYF 531
OY 335 GHAFVCDGYASDGTGFHFMWGVANGFYKLLLSPTSIGIGEGIGFTIYOEIITG--391
DB 532 DHIRNID---DGAALVHNHNTASN-----ITITGE-----SLTIDPT 567
OY 392 IEPKTPAEAGTDALPILALDIEAEYKSEGLNV-GYSIY-----NNGE 436
DB 568 ITPYNIDAPDEDNPPAFRRIRKDGGLY---LNLENTYALAKRGASTRSELPKNSES 622
OY 437 QSN-LDLCYRLKAD-----GEVEVKTSSINISWYGEHSEFSEFS 476
DB 623 NENMLYMKRTSDEKARNYNNHINERNANGFNGYGEEGKNNGNLNVTFKSKSEONREL-681
OY 477 LAPNLSOGINTITLLYRTGT--EOMEVPRHAOGYVNSIKVNTDP-----NNVYVLYVD 530
DB 682 ---LNGCTHNLNGDLYEKTLPLSGFTPHAD--IAGISSTKKDPHFAENNEVYVED 734
OY 531 ---NNEGKLSIYVNSFVADLNS-----YE-HSTITVVO-----558
DB 735 DWINRNEKATYMTVGNASISYSGRVANITSNITASKAQVHIQYKKGDTVCVASHDTGY 794
OY 559 ---FNSDSDPEIRTPVAFALSTGATADYISLGWMAEYPPGS-----S 599
DB 795 VTCITDKLSKALNSFNPTNLRGVNLTESANFYLAGRANLEGTIOSR--GNSQVRLTENS 852

```

Query Match 2.8%; Score 124; DB 9; Length 1541;
Best Local Similarity 20.4%; Pred. No. 1.8; Indels 322; Gaps 53;
Matches 197; Conservative 106; Mismatches

57 YVYRGDAERGTSOEE-----GSPATFYVNRGNNE--GYALVAAD 96
210 FIKKGDNTSLINNEHVGNNLKLVDATYGIAGTP---YVNHENNLIGFGNSKEE 266
97 DRFT-ILAVSPGRDM--DSMPDNLRMLOIYDOETG--LILSG----- 137
267 HSDPKILSODPLTNVAVLDSOSP-----LFYDREKGMLEFGSYDFWAGYKKSQOE 321
138 -----KAOLNEILRTGEPVPAEVALMDNGHFANDPMWNGYPM--NKKPELLPNGHAY 191
322 WNIYKQFTYDVJNKSAGSLISKTD-----YSSNGKSTYTGGKSL 367
192 TGCVAATAAQIMRYHSMPLGEG-----SFYHAGSLV-----GNMSG--- 229
368 NVDLADGDKPRNGKSVTFEGSGLTLNNHIDOGAGSLFPEGDEYVKGSTDNTPWAGAV 427
230 --TFGEMYDWINPGRPDLDNLTSQOVDAVATL-----MRDVSASVMSFEYENG 276
428 SVAEGTWTW--KVNHPOYDLRAKI--GKGLIVEGTGDNKSLKVGDTVLKQQTNG 482
277 SGTYYVYVYVGAAL--RNNFRYKRSLOLHVRLALYTSQEWMDIRGELASGRPVYVYAGNNO 334
483 SGCHAFASVGIYSGRSTLVANDKQVDPNISYIFG-----FRG---GR-LDLNGNSLTF 531
335 GAHAFVCDGASDGTFFHFMWNGVSNFGYKLLLSPTSLGIGEGIGFTYIOELTG--- 391

Query Match 2.8%; Score 124; DB 9; Length 1541;
Best Local Similarity 20.4%; Pred. No. 1.8; Indels 322; Gaps 53;
Matches 197; Conservative 106; Mismatches

57 YVYRGDAERGTSOEE-----GSPATFYVNRGNNE--GYALVAAD 96
210 FIKKGDNTSLINNEHVGNNLKLVDATYGIAGTP---YVNHENNLIGFGNSKEE 266
97 DRFT-ILAVSPGRDM--DSMPDNLRMLOIYDOETG--LILSG----- 137
267 HSDPKILSODPLTNVAVLDSOSP-----LFYDREKGMLEFGSYDFWAGYKKSQOE 321
138 -----KAOLNEILRTGEPVPAEVALMDNGHFANDPMWNGYPM--NKKPELLPNGHAY 191
322 WNIYKQFTYDVJNKSAGSLISKTD-----YSSNGKSTYTGGKSL 367
192 TGCVAATAAQIMRYHSMPLGEG-----SFYHAGSLV-----GNMSG--- 229
368 NVDLADGDKPRNGKSVTFEGSGLTLNNHIDOGAGSLFPEGDEYVKGSTDNTPWAGAV 427
230 --TFGEMYDWINPGRPDLDNLTSQOVDAVATL-----MRDVSASVMSFEYENG 276
428 SVAEGTWTW--KVNHPOYDLRAKI--GKGLIVEGTGDNKSLKVGDTVLKQQTNG 482
277 SGTYYVYVYVGAAL--RNNFRYKRSLOLHVRLALYTSQEWMDIRGELASGRPVYVYAGNNO 334
483 SGCHAFASVGIYSGRSTLVANDKQVDPNISYIFG-----FRG---GR-LDLNGNSLTF 531
335 GAHAFVCDGASDGTFFHFMWNGVSNFGYKLLLSPTSLGIGEGIGFTYIOELTG--- 391

Query Match 2.8%; Score 122.5; DB 10; Length 833;
Best Local Similarity 19.2%; Pred. No. 0.92;
Matches 142; Conservative 96; Mismatches 301; Indels 201; Gaps 35;

213 EGSFDY--HAGSLVGNMSGTFG--EMTD-----WINPGRPDLDNLTSQOVDAV 258
15 EGSINVLVDKAITGKPDGYGPTESIDRASAAVIFTKIINLPVD--ENAPFSFKDAKN 71

Query Match 2.8%; Score 122.5; DB 10; Length 833;
Best Local Similarity 19.2%; Pred. No. 0.92;
Matches 142; Conservative 96; Mismatches 301; Indels 201; Gaps 35;

213 EGSFDY--HAGSLVGNMSGTFG--EMTD-----WINPGRPDLDNLTSQOVDAV 258
15 EGSINVLVDKAITGKPDGYGPTESIDRASAAVIFTKIINLPVD--ENAPFSFKDAKN 71

QY 259 TLMRDVSASVSMSEFYENGSGTYSVYVGAAL--RNNFRYKRSLLQHLVALYTSOEHMDIRG 317
 Db 72 IWSSKXYIAAVEKAGVAVGDKENEFPEGKIDRASF-----ASMLVSAYNLKDKRVNG 122
 QY 318 ELAS-----GRPVYAGN---NOSIGHA-----FV----- 339
 Db 123 ELVTFEEDLLDHNGEERANILLINGISVGTGKWEPRKSVSRAEAMOFIALTRKKYKRD 182
 QY 340 -CDGASADGTFFHFMGWSNGFYKLLTSLGSGEGIGFTTYQELITGIEPAKTP 398
 Db 183 NAQAYVTD-----VKVSEPTKLTITGTGLDKLSADVT--LEGKAY 222
 QY 399 A-EAGTAPALPLAL-----KDIEAEVSESGL--NVGSIYNTGEGOSNLD----- 441
 Db 223 AIEASTDGTSAVYTLGKVAKNKDLTVYKKNOSFVTFYEVKKLAVERKLTFFDDDRAGQA 282
 QY 442 LGYRLNKADGEVILEVKTSSINISWYGEHPESFSLAPNOLSGINTITLLYRTGTEQW 501
 Db 283 IAKKLDERG-----NADVEYLNLANHDVRF--VANNLDGSPANIEFGEGEATSTTGK 332
 QY 502 EPVYRHAOGGYVNSIKV---NTTDPNNVVTVYDNNEGKLSIVPNSFVADLNSYEHSTIYV 557
 Db 333 LAVGIKOGDYKVEQVTKRGGLTVSNTGITTVKLNLDTPASAIKNVFA--LDADNDGVYV 391
 QY 558 QFNSDSDPELRTVYAFAL-STGATADDVYSLGVMAEVRG-----GSSNTPVY 604
 Db 392 GSKLSGRD-----PALNSQNLVVEGKASLKLKLVATIGEDKRVDPGSIKSNHGI 444
 QY 605 -----MSKDYLTISEGDTLWYRFSINNODEMCKISGSVAKTPTEVTHLFEV 653
 Db 445 SVVNNITAEAGAEALITIVGVDTYKDVYKFTTDS---RKL--VSKANPD---KLOY 495
 QY 654 GHNOTSTYTL-----DMAHNRVLPDETLKLNGLPENGELVYVFRTOSSSGS 700
 Db 496 VQKTLPTVEFTVTDQYGDPEFAGNTAIAKEVLP---KTGVVAEGGLDV---TDDSG 546
 QY 701 LW-----AAQETVHKQGE---TFVYKPYVEGPIPDGSIYATLHAFTNGQO--- 743
 Db 547 IGTGKTGVGNDVGEVTHQNGNATLSLKYVNTGEGVAFKFNELVSKVQVGGSPPT 606
 QY 744 QLYLKGRNYTVKIVGTAVEAIESESEIRFPNPRADYVEISAPCIPOETSIILEDLG 803
 Db 607 KIDLAVSTVEYQLSTYSDRYVSDPENLEGYEVESKNLAVADAKIVGNKV-VYTGKTPG 665
 QY 804 KI---VMAKNSLSAGHRMDV 820
 Db 666 KVDIHLTKGATAGKATVEI 685
 LT 19
 Db 09-839-894-6
 Sequence 6, Application US/09839894
 Patent No. US2002017686A1
 GENERAL INFORMATION:
 APPLICANT: Alboum, Zeev
 APPLICANT: Barry, Eileen M.
 APPLICANT: Levine, Myron M.
 APPLICANT: University of Maryland
 TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
 FILE REFERENCE: UOFPD.006A
 CURRENT APPLICATION NUMBER: US/09/839,894
 CURRENT FILING DATE: 2001-04-20
 PRIOR APPLICATION NUMBER: 60/198,626
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 6
 LENGTH: 867
 TYPE: PRT
 ORGANISM: E. COLI
 US-09-839-894-6

Query Match 2.8%; Score 122.5; DB 9; Length 867;
 Best Local Similarity 19.1%; Pred. No. 0.97;
 Matches 149; Conservative 105; Mismatches 238; Indels 287; Gaps 39;
 QY 43 RMGQTAVSDKISIDYVY-----ROGAERGITSQEE--GSPAYFYVANRG----- 85
 Db 230 RMDRTDLSQISGNFNFNLLPLPDIDGIRGTQSYIKNTDKRIAPVYVMLNFRSVERA 289
 QY 86 --NNEGVALVAADRPTI-LASPIGRDMDSPONLKMLOIYQOELIGLISGRAQIN 142
 Db 290 FRNNOLLGVWYLDGVSNEIDTARLPYGSYDLK-----LKIFE-----N 327
 QY 143 EEILRTGVAEYHAALMDNGHFANDPMRNK---OGYPNNKEPLPFGNHAYTGCVATA 198
 Db 328 TQVREELIP-----FNGRSGISGMQNDVFIQGNIIINDRYIERKNNKSSVNAEL 381
 QY 199 AAOIMRYHSPLOG-----EGSFYHAGSLVGNNSGTFGGEYDMINPFGNPDLDN 248
 Db 382 RLPTIRNIS-VQOGASVIDNKRYEGSLKWNNSGILSGSINSEFSPLMG-DNAKGNVQ--- 436
 QY 249 LQSGVDVATLMDVDSASVSMSEFYENG-----SGTYSVYVGAALRNFRY 294
 Db 437 -SISTYDGF-----SLSEYHNDKRVDCGNVYAGNSGCEYEST----- 473
 QY 295 KRSIQLVHVALYTSOEHMDIRGELASGRPVYVYAGNNOISIGHAFVCDGYASDGTFF--HFN 352
 Db 474 --SASLSIFPL-----GWTSTLIGTS---DIY-SSSYKNNHL 504
 QY 353 WKGWGSNGFYK-----LTLSPYSLGIGEGIGFTTYQEL-ITGIEPAKTPAEGTAL 406
 Db 505 SEGYFNQNIYKRGROQWOLTSSTSL---KMWYDFNFMALIGYNSQQLTDKGGYISV 560
 QY 407 PILAKIDBAEKSESGLNVGYSI--YNTGEOSN--LDLGYRLNKADGEVILEV----- 457
 Db 561 TL-----TRASRENSLAGSYNSRKYSNLEFVD-GWTSYNNNDYEHVRRRENK 612
 QY 458 -----TSSINTSWYGYGHPESFSLAPNOLSGINTITLLYRTGTEQWEPVYRHAQ 508
 Db 613 NRHNAGRLSGRIN--NFGDLNCSFSNNKR-----NTNSNNSLIGVGYSSPALSD 664
 QY 509 GGY-----VNSIVVNTTDPNNVVYV-----VDNNEGLK---SIVPNS 542
 Db 665 GFYWGSGASGLTKLAGGILIKVSNDKKNLVYVTALYGDVSLGSDNNAFIPVPAALTPAS 724
 QY 543 FVADLNSYEHSTIYV--QFNSDS-----PDEIRTPVAFAL-----STGATADYV 584
 Db 725 LIEDNNNGDKNISVLAPTNMDFILPGNVYVEIETKYSVSYIGRGFDKNGTPLSGAHV 784
 QY 585 ISLGVMAEVRPGSSNYPVWMSKDVLTLESGDYTLWYRFSINNODEMCKISGSVAKTP 644
 Db 785 INEPVHIDEDGFGS-----F 800
 QY 645 EYTHLFEVGHNOT-----STYLLDMAHNRVLPDETLKLNGLPFNSELVYVFROTOS 696
 Db 801 EYV-----GNKXTLFLKGRITTYCQLGKNKV-----HKGIVGVADVICDVNSTSS 846
 RESULT 20
 US-09-769-787-2
 Sequence 2, Application US/09769787
 Publication No. US20030091577A1
 GENERAL INFORMATION:
 APPLICANT: Microbial Technics Limited
 APPLICANT: Gilbert, Christophe FG
 APPLICANT: Hansbro, Philip M
 TITLE OF INVENTION: Proteins
 FILE REFERENCE: PWC/P21128WO
 CURRENT APPLICATION NUMBER: US/09/769,787
 CURRENT FILING DATE: 2001-01-26
 PRIOR APPLICATION NUMBER: GB 9816337.1
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: US 60/125164
 PRIOR FILING DATE: 1999-03-19

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:24:27 ; Search time 26 Seconds

(without alignments)
3116.974 Million cell updates/sec

Title: US-10-030-330-1

Perfect score: 4438
Sequence: 1 MKKSFLLAIYVLEGIAMQGH.....PNCAYILKVDGYTTKINIVH 843Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 283224 seqs, 96134422 residues

1 number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:***
2: p1r2:***
3: p1r3:***
4: p1r4:***

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1196	26.9	886	2 A10890	cysteine proteinase
2	494	11.1	399	2 A37768	streptococcal pyro
3	388	8.7	337	1 BPSOP	streptopain (EC 3.
4	179	4.0	939	2 I41197	ase protein (enter
5	173	3.9	3624	2 AD0835	large repetitive p
6	163.5	3.7	1155	2 H71456	probable pyrolysin
7	157.5	3.5	4936	2 AH2515	hypothetical prote
8	154	3.5	1325	2 A64905	yeak protein - Esc
9	149.5	3.4	1471	2 D72758	hypothetical prote
10	146.5	3.3	1343	2 E90893	hypothetical prote
11	145	3.3	918	2 T02759	hypothetical prote
12	144.5	3.3	1256	1 A43829	mutamidase-release
13	144.5	3.3	1343	2 D85724	hypothetical prote
14	143	3.2	3263	2 E82410	hypothetical prote
15	142.5	3.2	940	2 AD1374	internalin protein
16	142.5	3.2	1349	2 A11476	cell surface prote
17	142	3.2	682	2 J00420	beta-1,3-glucanase
18	141.5	3.2	827	2 T20492	hypothetical prote
19	141.5	3.2	1308	2 A90428	hypothetical prote
20	141.5	3.2	1771	2 S76851	serine proteinase,
21	141	3.2	709	2 G64057	hypothetical prote
22	140	3.2	1348	2 AH1115	IDA-specific metal
23	140	3.2	1461	2 E90696	cell surface prote
24	140	3.2	1461	2 A85547	hypothetical prote
25	139.5	3.1	814	1 KXBY	hypothetical prote
26	139.5	3.1	890	2 A30461	kekin (EC 3.4.21.6
27	139	3.1	1417	2 A90670	bacteriophage BCN5 -
28	139	3.1	1417	2 D85521	probable invasins
29	139	3.1	4199	2 S76412	probable adhesin e
					hypothetical prote

30	138.5	3.1	650	2 S48751	major surface prot
31	138.5	3.1	2902	2 C71953	toxin-like outer m
32	138	3.1	1879	2 S74915	extracellular nucl
33	138	3.1	3283	2 AC1018	large repetitive p
34	137.5	3.1	647	2 A49218	hemagglutinin homo
35	137.5	3.1	649	2 S48752	major surface prot
36	137	3.1	754	2 T14877	hypothetical prote
37	137	3.1	1310	2 AE1749	glycosidase homo
38	137	3.1	1398	2 T28159	pyrolysin (EC 3.4.
39	136.5	3.1	2236	2 AD1018	large repetitive p
40	136.5	3.1	4436	2 E71086	hypothetical prote
41	136	3.1	5188	2 B85547	probable RTX fami
42	136	3.1	5291	2 F90696	hypothetical prote
43	135	3.0	987	2 A64474	hypothetical prote
44	135	3.0	1829	2 AE1864	hypothetical prote
45	135	3.0	1939	2 D97316	probable S-layer p

ALIGNMENTS

RESULT 1

T10890 Cysteine proteinase (EC 3.4.22.-) / hemagglutinin protein - Porphyromonas gingivalis

C:Species: Porphyromonas gingivalis

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999

C:Accession: T10890

R:Madson, T.E.; Clark, V.L.; Kuramitsu, H.K.

J Infect. Immun. 63:238-247, 1995

A:Title: Revised sequence of the Porphyromonas gingivalis prt cysteine protease/hema

A:Reference number: 217199; MUID:95105001; PMID:7806362

A:Accession: T10890

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-886. <MAD>

A:Cross-references: EMBL:S75942; MID:9913136; PID:9913137

C:Genetics:

A:Gene: prt

C:Keywords: cysteine proteinase; hemagglutinin; hydrolase

Query Match	Score	DB 2	Length	886
Best Local Similarity	26.9%			
Matches	290	Conservative	155	Mismatches 340; Indels 70; Gaps 23;
QY	1	1 MKKSFLLAIYVLEGIAMQGHSAPTKRALSLAR--LALQVSLRMQOTAVSKISTDIY	58	
DB	19	MKRIFTYLGLLCLPML-QAGPYTRSKAEQTAFFAKRQPTL-----SSSTASLRMDEV	73	
QY	59	YRQGDARQITSGEGSPAYFYVANRGNNEGVALVAADDRPTILAYSPIGRFMDSPD	118	
DB	74	YKAAREE-----ALFEVNRKEKDFLLVADDRPEVIGYAFKCHFDAAAMPD	123	
QY	119	NLRKMLQIYDOETGLILSGRAQNEELR--TEGVPAEYHALDNGHFAHPKRNQY	175	
DB	124	NLRGWLKGYEREMLAVDKRAEPIPIREKPTNDLPSIAPIETGEHMSDPLMGGY	183	
QY	176	PWANKPELLPNGNHAYCYATAAOTMRHSPVLOGEGSFYAGSLVG---NMSGTFG	232	
DB	184	PFNTLHPLPSGQOAYCVATAMGQIMRHVKPEKASGEVDY--DMTGTHHYSGTFG	242	
QY	233	EMVDIMPGNPDNDNLTQSOVAYATLMDVSAVSMSFENGSGTYSYVVGALNNF	292	
DB	243	ETIWMKMPGNISY-GISPEVYKALSTFMRVSVSYMOPADFSGSTFVEPALKEET	301	
QY	293	RYKSLQHLVRAVLTSEWMDIRGELASGRPVYVYAGNNSIGHAFCVDSYASDGFHFN	352	
DB	302	HYKSLVYIHRSLPGLKEMKMDIKELAEKRPVYVYAADSMGHAFCVDSYASDGFHFN	361	
QY	353	WNGGVSNGFYKLLTSLPSLGTGEGSIGTYIOETITGEPAPKPAEAGTALPILAK	412	
DB	362	WNGGVSNGFYKLLTSLPSLGTGEGSIGTYIOETITGEPAPKPAEAGTALPILAK	420	
QY	413	DIEAEVYSEGLANVGYSTYNTGEGSNLMDGLYRANKADGEVIEKTSINISWGC-YGEH	471	

[illegible]

RESULT 2
A37768
streptococcal pyrogenic exotoxin type B precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
C:Date: 31-May-1991 #sequence-revision 31-May-1991 #text-change 22-Jun-1999
C:Accession: A37768
R:Hauser, A.R.; Schlievert, P.M.
J. Bacteriol. 172, 4536-4542, 1990
A:Title: Nucleotide sequence of the streptococcal pyrogenic exotoxin type B gene and related
A:Reference number: A37768; MUID:90330563; PMID:2158264
A:Accession: A37768
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <HAU>
A:Cross-references: GB:M66905; GB:M35110; NID:q153818; PIDN:AAA26978.1; PID:q153819
Keywords: streptococcal cysteine proteinase
Keywords: exotoxin

[illegible]

OY 27 ENSGRTSYVYVGCALNNRRYKRSLQHLHRAALYTSQEHNDIMRGLASGRPVYVYAGNNS 333
 Db 278 PSSGSGAGSSSRVQALKEENFYQNSVHQINRSPFSKQDEAOIDELSONQVYVYQGVGV 337
 OY 334 IGHAFVODGYASDGTFFHFNMGNGVSNGFYKLTLLSPTSLGIGEGIGFTYIOETITGIE 393
 Db 338 GGHAFVYIDGADGKNFFHYHVMGNGVSDGFEFLDALNPSALGTGGGAGGENGYSQSAVVGIK 397
 OY 394 P 394
 Db 398 P 398

RESULT 3

streptolysin (EC 3.4.22.10) precursor - Streptococcus pyogenes
N:Alternate names: streptococcal cysteine proteinase; streptococcal peptidase A
C:Species: Streptococcus pyogenes
C:Date: 24-Apr-1994 #sequence_revision 07-Oct-1994 #text_change 07-Oct-1994
C:Accession: S07668; A00978
R:Yonaha, K.; Elliott, S.D.; Liu, T.Y.
J. Protein Chem. 1, 317-334, 1982
A:Title: Primary structure of zymogen of streptococcal proteinase.
A:Reference number: S07668
A:Accession: S07668
A:Molecule type: Protein
A:Residues: 1-337 <YON>
R:Tal, J.Y.; Kortt, A.A.; Liu, T.Y.; Elliott, S.D.
J. Biol. Chem. 251, 1955-1959, 1976
A:Title: Primary structure of streptococcal proteinase. III. Isolation of cyanogen bromide
A:Reference number: A00978; MUID:76190087; PMID:1270417
A:Accession: A00978
A:Molecule type: Protein
A:Residues: 85-107,'L',109-244,'N',246-337 <TAI>
A:Note: Trp-298 is at the binding site of the enzyme
C:Comment: This enzyme can remove the activation peptide from the proenzyme.
C:Superfamily: streptococcal cysteine proteinase
C:Keywords: cysteine proteinase; hydrolase; zymogen
F:1-84/Domain: activation peptide #status experimental <PRO>
F:85-337/Product: streptolysin #status experimental <MAT>
F:131,279/Active site: Cys, His #status experimental

Query Match	8.7%	Score 388;	DB 1;	Length 337;
Best Local Similarity	27.8%	Pred. No. 1.4e-16;		
Matches 108; Conservative	63;	Mismatches 138;	Indels 80;	Gaps 12

```

QY      26 KEKALSLARLAROYLSRBYGOTAVASDKTSLIDYVRNQGSAEGLTIOBEGSGSAFVYUANRG 85
Db      9 KEKDSALITFIQKSAAIKARGASAE-ITKIDVNLGGEL-----SGSMIYVNIASG 59
QY      86 NNEGVALVAADDRIPTITLTAISPGRFDDMSMDLRLMWLOIYDQIGLLSGRAQNLNEET 145
Db      60 -----MESUYEQIK-----ENKLDITTYAGTETIKQPV 87
QY      146 LRTEGVPAEYHALMDONGHAMPMBRMNGGYPMNNKEPLRP---GNHAYTG-----CVA 196
Db      88 -----VKSLLD-----SKGIHYNOGNPYNLTTPYIEKYKRGEDOSFVGQAATGICVA 133
QY      197 TAAAOITMRHYSPMLQEGGSFDY-----HAGSLVGNMSGTGFEYDMINNPGNBDL 246
Db      134 TATAQIMTKYHNPDKLKNYVTLTSSNDPYEDHPKNTL---FAALSTRQYDMNNI--LPTY 188
QY      247 DNITGSOVD-AATITLMRDVSASVSMSPFENGSGITSYIVYGALFNNNRYKRSIQLHNRAL 305
Db      189 SGRQSOQNVKMAISELMDADVGLISVDMDYPPSSGSAGSSHYORALKENGVNOGYOJDRGD 248
QY      306 YTSQEHMDIRGELASGRPYVYAGNNOISIGHAFVODGASPDGTFHFMNMGAGSVNGFYKL 365
Db      249 FSNQMDENADIDLEISQNPVYVEGYGKXGGAFAVIDGAGNRFHYNDMGNGVSDGFFRL 308
QY      366 TLISPTSLGIGEGIGFTIYOEITITGIEP 394

```

Db 309 DALNPSALGTGGAGGNGEYSAVVGIKP 337

RESULT 4

141197 eae protein (enteropathogenic) - Escherichia coli

C/Species: Escherichia coli

C/Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 26-Aug-1999

C/Accession: I41197

R/Jerse, A.E.; Yu, J.; Tall, B.D.; Kaper, J.B.

Proc. Natl. Acad. Sci. U.S.A. 87, 7839-7843, 1990

A/Title: A genetic locus of enteropathogenic Escherichia coli necessary for the production

A/Reference number: I41196; MID:91045893; PMID:2172966

A/Accession: I41197

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: DNA

A/Residues: 1-939 <RES>

A/Cross-references: GB:M58154; MID:g145852; PIDN:AAA62775.1; PID:g145854

Superfamily: Invasin

Query Match 4.03; Score 179; DB 2; Length 939;

Best Local Similarity 20.38; Pred. No. 0.0057; Mismatches 330; Indels 254; Gaps 48;

Matches 179; Conservative 119; Mismatches 330; Indels 254; Gaps 48;

QY 9 IYVLFQIAGHSAFVTK-----ERALSRLALROVSLMGQTAVSDKISID 56
Db 132 LVAAGVA--GHTNKLTKMSPDYTKSMDDKALNTA---AQAASLSGQSLSRLNGD 185
QY 57 YVYRQDAERGITSQEGSPAYFYVANRG-----NNEGVALVAADDRIP---T 101
Db 186 YA---KDTALGIAGNQASQQLQMLQHYGTAEVNLQSGNPFSSSL---DFLLPFYDSEK 239
QY 102 ILAYSPIGRDMDS-----MPDNLRMLQIYQDEIG---LILSG----- 137
Db 240 MLAFGGVARYIDSRFTANLACORFLLPENMLGYNFIDQDPSGDNTRLGIGGEYRDY 299
QY 138 -KAOLNEELIRTEG-----VPAEVHALMDNGHNPND---KRMNGYPMN- 178
Db 300 FRKSVN-GYFRMGSMHSTYKKYDERPANGFDIRFGYLPSPALGAKLMEDYIDNV 356
QY 179 ---NKEPLPENGHAATGCVATAAQAQIMRYHSMPLQEGSGFDYHAGSLVGN---WSGT 230
Db 359 ALFNSPKLOSNPCAATVYG-----VNTPIPLVYMG-IDYRHGT--GNENDLLYSMQ 406
QY 231 FGEWYD--WINMGNPDLDLQSOVDAYATLMRDVSAVSMSFEYENG-----GTY 280
Db 407 FRYQDFPW-----SQIEFOYVNEELTKTSGSRVLDVQRNNNIILEYKODIL 454
QY 281 SYVYVGAALRNFRKRSIQHLVRLALYTSQ--MHD-MIRGELASGRPVYAGNOSIGHA 337
Db 455 SLNIPHDINTERSTOKIQIIVSKYGLDRIYMDLSLRQ---GGQIOHSGSGSADYQ 511
QY 338 FVCDGIASDGTFRHNMGMGVSNGEYKLLTLLSPTSLIGEGE-----GFTIYOET 388
Db 512 AILPAYVOG-----SNVYKVTARAYDRNGSNNNVLLITVLSNCOVDQV 559
QY 389 ITGIEPAKTPAEA-GTDALPILALKDIEAYKSSGSLNAGVSYIYNGEBSNLDIGRLN 447
Db 560 VTDPFADKISAKADGDEAITTYTA-----TYKKNGVAQANPVSP--N 599
QY 448 KADGEVLEVTSSTINISWYG-----YGEHPES--FSLAPNQLSQGINTTLLY---- 493
Db 600 IVSGTAV-LSANSANTGSGKATVTLKSDKPGQVYVSAKTAEMTSALNANAIVFVDTKA 658
QY 494 --RRTGTEQWEPVRAHAGGVYNSIKVNTDP--NNVVYVVDNNEKSLIYVNSEFVADLNS 549
Db 659 SITEIKADKTTAAVANGODAITTYVYKWKGDKPVSNQVETFTTLIGKLS--NSTEKTDTNG 716
QY 550 YEHSTIVQEN-----SDSPDEIRTP-VAFALSTGATADVDVLSGLVMAEVPBGSS 599
Db 717 YANVTLLSTPGKSLVSARVSDVAVDKAEVEE--FTTLTID--GNIEIVGTGVK 770
QY 600 NYEVVMSKDVLTLSGDDYTLMYRFSINQNDKWKI-----GSVSVKTPREYTH 648

Db 771 KLPTW-----LQYGVNL--KASGNGKRYTRSANPALASVDASSGQVTLKEKGTITI 822

QY 649 PLEFVGHQSTSTYTLDMANRVLPDF-----TLKNLG--LPFN-GEVLYVFR-- 652

Db 823 SVIS-SDNOTATYTIATIPNSLIVPMNSKRYTYNDVAVTCNFGKLPSSONLEENFKAW 881

QY 693 -----OTGSSGSL--WMAOETVHIKQGEFTFYKPPVEGPI 726

Db 882 GAANKTEYKSSOTITISWQQTADAKSGVASTYDLVKONPL 923

RESULT 5

AD0835

large repetitive protein [imported] - Salmonella enterica subsp. enterica serovar Typh

C/Species: Salmonella enterica subsp. enterica serovar Typh

A/Note: this species has also been called Salmonella typhi

C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C/Accession: AD0835

R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moute, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A/Reference number: AB0502; PMID:11677608

A/Accession: AD0835

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-3624 <PAR>

A/Cross-references: GB:AL513382; PIDN:CAD05867.1; PID:g16503842; GSPDB:GN00176

A/Genes: STR2875

Query Match 3.98; Score 173; DB 2; Length 3624;

Best Local Similarity 20.38; Pred. No. 0.11;

Matches 223; Conservative 102; Mismatches 376; Indels 400; Gaps 54;

QY 14 GIMQGSAPVTERALSLRLALROVSLMGQTAVSDKISIDYVYQSGAENGIT----- 69
Db 2545 GIA-AGFTADPTREAPRITNVVDVGIIYTAIANGCVTNDQPTLNGTAQAGATVSIY 2603
QY 70 -----SQEGSPAYFYVANRGNEGVALVAADDRIPITLAYS 106
Db 2604 NNGALGTTTANNSGNNSFTPTGNTLREGSHAFATATANNAGTGSVSAATVYDTLAPGT 2663
QY 107 PIGRFMDSPMDLRMLQIYDDEIGLISGKAOLNEELRT--EGVPAEVHALMDNGHF 164
Db 2664 PSLTSLAD-----GSLSGOAEANSTVYVTLAGVTLTTTA-GSNG-- 2703
QY 165 ANDPMRNGQYPMNKE-----PLP--NGMHAATGCATA 198
Db 2704 ---AWSLTLPRKQLEGQLINVTANDAGNASGCTGATPILPLAARDNITSIDLTSTA 2758
QY 199 AAOIMRYHSMPLQEGSGFDYHAGSLVGN-----WSGTGEM----- 234
Db 2759 VTSQNTQSDGILLVGLG--NVASVGLNDIAQVETITAEGBGTGVITIDAATGIVLSLS 2817
QY 235 -----YD-----WINMGNPDLDLQSOVDAYATLMRDVSAVSMSFEYENGSTYS 281
Db 2818 TOEIVQRYDTSIGTW-----TTIVTAVGDDEFANLLTLGSGVTLNLNGLAGEQYR 2868
QY 282 VYVVGALRNFRKRSIQHLVRLALYTS--QEMHDMINGELASGRPVYAA-- 330
Db 2869 V-----LTNTSL--ATGSYSLVDVQTSAG-TISG-PTISTGNVMAADTAPT 2915
QY 331 -----NOSIGHAFVCDGYASDGTFRHNMGMGVSNGEYKLLTLLSPTSLIGIG- 376
Db 2916 GTTATATNANGSTPAGAGV-DILGQYGTILHN-----QGSYITTLTKRTPA-GYGH 2967
QY 377 GEGIGFTIYQ-----EITGIEPAKTPAEA-GTDALPILALKDIEAY-----KSES 422
Db 2968 KESFTYITINGVSSAAQLVINLGPAPVPGSVIATDNNSLIV--DTHVSYVNNGSTOS 3026

Db 705 GATVFNAGNTYSGKTLVNDGLITLTAAGTGMG--SSEVTIANPG-----TLD 753
OY 609 VL--TISEGDTLWYRS-----INNCKDEKKIGSVKPTETETHPLEFVGHNGSTYR 662
Db 754 ILASTASAGDYTLTNLAKGDKLKRVLSSDDKAEFGTHATGEFA---GVALCKDSTFP 809
OY 663 LDMAHNRVLPDFTLKNLGLPFNGELVYVFPOTQSSGSLMAOETVHIKOGETFFV---- 718
Db 810 LERONTAALTHMALQS-----DSENTTGVKVGESLIGGL-----ANNCGTIIIDTDI 856
OY 719 --KPVVEGPIPDOSYATILAFVNGOOLYLKGRNITYKIVNGTVEALIESEETRPV 776
Db 857 PAATLAEYI-----SVDTLVVAGADYTWKG-RNYQ---VNGTSDVDLID-----VP 898
OY 777 NPARD 781
Db 899 KPMND 903

CT 9
58
Hypothetical protein APE0057 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72758
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450, MUID:99310359; PMID:10382966
A:Accession: D72758
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1471 <KAN>
A:Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BAA78966.1; PID:d1042742; PID:9510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0057

Query Match 3.4%; Score 149.5; DB 2; Length 1471;
Best Local Similarity 18.3%; Pred. No. 0.76;
Matches 192; Conservative 158; Mismatches 352; Indels 345; Gaps 48;

OY 13 FGIAMOGHSAPYTKERALSRLARLAKQVSLRMGQTAVSDKISIDYVYRQDAGERI--T 69
Db 41 YTIAYEGGAVIYAVAGNVAVASYSSEVELVTG---DGLATVYVYAREPRVYVSLXS 95
OY 70 SOEGSPAYFYUAN-RGNNEGVALVADDRIPITLAYSPIGRFDMDSMDNLRMLQITD 128
Db 96 GGEKVAQASFEIIOULSGESDPRPDIVDDSLAVSTLTSPPSGVEILLAPFIPSSD 155
OY 129 OEIGLILSGKALNEILRTGEVRAVNAIMDNGHAFANDPMWNOGYRPNKNEPILLPN 188
Db 156 GAAVYVUTTSIVYDRV---SISVDNTKREYGVADALLILASGVSPEKADITYQCG- 210
OY 189 HAYTGCVALA-AAQIKRYHSPLOGSGFDYHAGSLVGMWSTFGEMTDWIMNPGNDID 247
Db 211 -AVLEFKASGDSGFVKKLSFNLITDGVLEVRIGLGTG-----LELE 251
OY 248 NLTQSG-----VDAYATLMRDYSASVSMSFENGSGTYYVYUGALRRNFRKRSL 298
Db 253 YLTIDASGRKVLPLPSISL-GDALATGNTISLHNVEGFLNVVQVNSDPEKI----- 304
OY 299 QLVHVALYTSQEMHDMIRGELASGR---PVUYAGNNOSIGAFVCDGYASDGTFFHNMGM 355
Db 305 -LAVRLELSGRPMRGL---ELVSKHNHVRLY-----IYEGRASQMEFLRIRGV 348
OY 356 GGVS---NGFYKLTLSPLSLGIGGEGITTYOETITTEIRAKTPAEAGTDALPLAK 412
Db 349 GVLALEPSPILKLT--PLEVLEGG---SYFYKOLPLAAVTVASTNP--TEDPLDLAP 400
OY 413 DIEAEKYSBSGLNVG-SIYNTGEOSNLDLGR---LNKADGEVIEVKTSINISWVG 467

Db 401 SVEVSKSGSG---GVFAIAGITMLEADMSIGNESSLVLTALSYAKLDVDFAINSSTII 457
OY 468 YGEHPS-FSLAPNQSQINITTL-----LYRTGTQMEPVHQAQ-GYVN----- 513
Db 458 IGRPEGLPVAPPEADTILNTVDLNVAEAEIARDAIKADIEFVAGSGGITMGAVLA 517
OY 514 -----SIKVN-----TDPNNVY---TVDN-----EGKLSIVNSFYA 545
Db 518 EAMNLSKNGAVASATVASSILVGDIVLSNNIEAVSTTILTSPIETRISLGRNTSLT 577
OY 546 DLSNEHSTITVO-----FNSDPDEIRPVAPALSTGA----- 579
Db 578 TYSLSHSPLIYVSGTNGSSTFTGASVTKLGLFNLTPD-----LTIYSSVLELA 632
OY 580 -----TADVISLQWMAEVPGGSSNPVWMSKVDVTLSEGDYTL 619
Db 633 YKTRGSITLMREGKASIELTDPDSLT---LIALWGS-----ITLEDSPFK- 677
OY 620 WYRFSSINQKDE-----MKRI-----GSYS-----VKPTPEYH 648
Db 678 --KFTYKTRDDLTIFMKGVAVYERASFEVSYMSKLRKGVDELASASLIMTAPEGEI 735
OY 649 PL-----FEVGNH-----OTSTYTLDMANRVLDPFTLKNL-----GLPFNGE 686
Db 736 TLEDTLISGVNFRVDESEAMPYRETSNVYVEGSSNAI---TVKSVETIISTYINAT 792
OY 687 LVVFPNQOSSGSLMAOETHIKOGETFFVYKPVVEGIPDGSYRATLHAVNQ----- 742
Db 793 YTVIVSEIASIGSSITQGG-TLNTQASNLV-----LQDNTLNGEYNTLSSQTSV 843
OY 743 --QOYLKGRNITYKIVNGTAVEALIESEETRPVNPARDVEIASCPIDETSIILFD 800
Db 844 EVKRLTVECCSQITLDGVCG-----YVESASGGGSLALQFN 882
OY 801 -----LSGKITYKNSLSAGHRADYSRLPENGAYI----- 829
Db 883 GGVTYLTGEV-----EGELEYS-LSNGARLMSMLEVKGGLRIVGEGDVEGSA 932
OY 830 -----LKVQDYTKRINIY 842
Db 933 SLVVECIARASATSLSEVSGLTADLNIV 959

RESULT 10
E90893
Hypothetical protein Ecs2117 [imported] - Escherichia coli (strain 0157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: E90893
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gawawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90893
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1343 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA83540.1; PID:dl3361583; GSPDB:GN00154
A:Experimental source: strain 0157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: Ecs2117

Query Match 3.3%; Score 146.5; DB 2; Length 1343;
Best Local Similarity 21.7%; Pred. No. 1;
Matches 155; Conservative 84; Mismatches 244; Indels 231; Gaps 40;

OY 161 NGHFANDPMWNOGYRPNKNEPRL---PNCNHAYTGCVALAAQIMRYHSPLOGEGSFD 217
Db 295 NGYFNGAVNISNNGLLNNKREYSLGVQDGSNGVNVYTDK----- 334
OY 218 YHAGSLVGNMS--GTFGE--WYDWIMNPGNFDLNLDSQVDAIATLMRDYASVSMSFY 273


```

Db 335 -----GHNHFGT-GEARFYITIGAGGELVNSREGKD-----SGITIGMK 377
Qy 274 ENSGCTYV-----YVGCALNNERYKRSLOLHV--RALYTSQEHMDIRGELASGRPVY 327
Db 378 ETGNGNLTVKDKNSVITNLCTNLGYDGHGEMINISNEGLVYVNSGSSIGCTGVGKSYIT 437
Qy 328 AGNNOSIG-HAFVCDGVASDGTFFHNMG-----WGVSNGFYKLLTSTST----- 372
Db 438 TGGIWEVKNKVVYTTIGVAGVGNLNI SDGKFKVSONITFLDGKASGIGTLNLMDSSEFDT 497
Qy 373 -----LGIGEGEI-----GFTIYOE--IYGIIEPAKTPAEAGTALPILAKIDIEAYKS 420
Db 498 VGINVNGFSGGIYVNSGATLNSGCGFIGNASGKIIVNISTDSLNNKLTSTNNAL-- 555
Qy 421 ESGLANGYSIYNGEOSNLDLGRKLNKADGEVIEVKTST-----NISWYGEHP 472
Db 556 ---LVGV--VLGTGE-----LNITGGIYKADTQIALNDKSKGVRVDGQNSLL 600
Qy 473 ESFLAPNOLSGINTITLLYRRT---GTEQW---EPV-----RHAQGGYV 512
Db 601 ETENMAYG--TSGTGLITLNSGTLNVEGGEVYLGVEPAVGLNLGAHGEAAAGFI 658
Qy 513 -NSIKV-----NTDPPN--HYVYTVDNNEGKLSIVPNSFVADLNSYEST 554
Db 659 TNATKVEFGSGEGEYFENHTNNSDAGYQVMDLITGDDKDKVI-----HDA 704
Qy 555 ITVOFNSDPEDEIRTPAALSTGA--TADVISLGVMVAEY---PGSSNPVYVSKD 608
Db 705 GHTVFNAGNTYSGTILVNDLITLASHADGVTCMG--SSEVTLASPG-----TLD 753
Qy 609 VL--TISEGDTLWYRS-----INNOKEMKKIGSVKTPLETETHLFEVGHQSTST 662
Db 754 ILASTNSAGITTLNALKGGLKRVQLSSDPMKGFTHATGEPA---GAQLKSTST 809
Qy 663 LD-----MAHNRVLPD---TLKNLGLPENGELVYVROTOSSSGLMAAETVHIKOG 713
Db 810 LEBDNFALTHAMLOSDIENTVSNG-----EGSIGL-----AMNG 847
Qy 714 ETIVY-----KPVESPIPDGSRATLHAFVNGQOOLYLKGRNTYVKIYNGT 761
Db 848 GTLIFDTIDPAATLAEGYI-----SVDTLVYGASDYTWKG--RNYG--VNGT 890

```

RESULT 11

T02759
 hypothetical protein - Acinetobacter sp. (strain ADP1) (fragment)
 Species: Acinetobacter sp.
 Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
 Accession: T02759
 R:Segura, A.; Ornstom, N.L.
 Submitted to the EMBL data library, June 1997
 A:Description: P2R153.
 A:Reference number: Z14724
 A:Accession: T02759
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-918 <SEG>
 A:Cross-references: EMBL:AF011339; NID:g2286203; PIDN:AA027114.1; PID:g2286204
 A:Experimental source: strain ADP1

Query Match 3.38; Score 145; DB 2; Length 918;

Best Local Similarity 18.08; Pred. No. 0.68; Mismatches 282; Indels 344; Gaps 45;

Matches 163; Conservative 127; Mismatches 282; Indels 344; Gaps 45;
 Qy 94 AADRTPTILAYSPIGRFDMSMDNLRMLQIYDQIGLILSKAOLNEELITRECVPA 153
 Db 97 ATDPTPTLVGTGVAV-----VLSVDGGAIVGSAVVDSSNG--- 134
 Qy 154 EVHALMDNGHPANDPMKRNOCIPMNKKEPLLPGNHAHYTCVATAAQLRHSWPLQGE 213
 Db 135 -----NMSYTLPSQSEGTSHYSTGTASNAAG-----TQGT 163

```

Qy 214 GSPDIHAGSLVGNMSTFGEIMYIMNPANDLNLQ--SQVATATLMDVASYSMS 271
Db 164 ASFTLLIDTIDTAPD--VPISGVNDVDVGLIQP---LTQASTDPTPTPL----- 207
Qy 272 FYENGSGYSVYVVGALRNRRYKRSLOLHVRLALYTSQEHMDIRGELASGRPVYAGNN 331
Db 208 -----TGT-----GATPG-----DVAK-----VYQOND 225
Qy 332 QSIHAFVCDGVASDGTFFHNMGCVSNGEYKLLTSPISLG-IGEGIGFTIYOEIT 390
Db 226 -LVGSVTV-----GADGMNNTYIIPAGLREGSHDLSVATDPVNGESGSDPFTLVDLTA 280
Qy 391 GIEPAK-----PAEGTALPILAKD-----IEEYSESG-----LNV 426
Db 281 PPAEGTENGSGSEITSAEAGS---VVKVKGANGVLSGATADSSGKVTYTLADPLND 336
Qy 427 GYSIYNT-----GEE-----QSNLDLGRKLNKADGEVIEVKTST 461
Db 337 AETVKITATKAGNESQATDLTAPNVIIADENIVEAKVDPTTYVYVNPSETIINETSIL 396
Qy 462 NISWYGEHPESFSLAPNOLSGINTITL-----LYRRTGEQWEPVRAHQ 508
Db 397 NI---GSTYPGYTVGODQIADAVISVSTSLINLFDNAEMKLYKQODDSQWELIADNQ 453
Qy 509 G-----GYNSIRVN----- 518
Db 454 SPGLLDLIGFGQTTKYTAGELRPGNRYRDEFTGSLIGLSTIRADLQLTQNTAMPV 513
Qy 519 ---TTDPNNVYTVDNNEGKLSIVPNSFVADLNSY-----EHSITVQF----- 559
Db 514 GDITSKSGNVITTDADATNGQDVYSQRYKTVAVNGQVAAADGTTIVGEHGLTIKAGSY 573
Qy 560 ---NSD-----SPDEIRTPAALSTGA--TADVISLGVMVAEYPGSSNPVYVSKD 608
Db 574 KYTPNSVTVYAGKIDFTFNITTTDA-STGKSDPTAKLIQI-----GTNSDLDLTWPN 624
Qy 609 VLTISEGDYTLWYRFSINNOKEMKKIGSVKTPLETETHP-----LFEVGHQSTST 661
Db 625 ---NPEADAT---SVATNDE--DTVIGIGTNETTETTAGSINESWILIGGSGQTVYSQ 675
Qy 662 TLDMANHRV-----LPDTLKNLGLPENGELVYVROTOSSSGLMAAETVHIKOG 715
Db 676 TITIAQNLGAVEIGLSTSSLLG-----GAASVQFQKLVNNT---WTVDTIASSLAD 728
Qy 716 FYKPVVEGPIPDGSYTA--TLHAFVNGQOOLYLKGRNTYVKIYNGTAVPAISSEIR 773
Db 729 LI-----GLFPNGTGVYDLEA---GQYRYIILNNG--LGVAGNVSSEVTSDDL 777
Qy 774 VEPAPADVEIAPCIQIETSIILFPLDSGKIVK-----NLSAGGRMDVSR----- 823
Db 778 SYTTSREIVQ-----GNVLTDTGTAGVAKVASHYTVDAISHDGLTYTVTSTGI 827
Qy 824 ---PNCAYILKVDG-YT 836
Db 828 SILGTHGTVIKSDGSYT 845

```

RESULT 12

A43829
 muramidase-released protein precursor - Streptococcus suis (type 2, strain D282)
 C:Species: Streptococcus suis
 C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
 C:Accession: A43829; S21549
 R:Smith, H.E.; Vecht, U.; Gielkens, A.L.; Smits, M.A.
 Infect. Immun. 60, 2361-2367, 1992
 A:Title: Cloning and nucleotide sequence of the gene encoding the 136-kDa dalton surf
 A:Reference number: A43829; MUID:92267650; PMID:1587602
 A:Accession: A43829
 A:Molecule type: DNA
 A:Residues: 1-1256 <SMI>
 A:Cross-references: EMBL:X64450; NID:g47550; PIDN:CAA5781.1; PID:g47551
 C:Superfamily: muramidase-released protein

C keywords: tandem repeat; transmembrane protein
 F:1-47/Domain: signal sequence #status predicted <SIG>
 F:48-1256/Product: muramidase-released protein #status predicted <MAT>
 F:663-681/Domain: small repeat <RP1>
 F:839-861/Domain: small repeat <RP2>
 F:953-1006/Domain: large repeat <RP1>
 F:1064-1084/Domain: large repeat <RP3>
 F:1089-1142/Domain: large repeat <RP3>
 F:1143-1195/Domain: large repeat <RP3>
 F:1232-1248/Domain: transmembrane #status predicted <TM1>

Query Match 3.3%; Score 144.5; DB 1; Length 1356;
 Best Local Similarity 18.7%; Pred. No. 1.2;
 Matches 214; Conservative 121; Mismatches 374; Indels 433; Gaps 60;

```

21 SAPYKERALSLARALROVSLR-----NGOTAVSOKISIDYVROGDAREGTTG----- 70
83 SAVATTSEVEKAAVALEQVTSSEPLAGLQKELAKTEDATLAKAIEDAQTAKLAAR 142
71 -----OEGSPAYFYVANNR--GNN-EGYA---LVAAADRIPITLAVSPIGRFMD 114
143 LADESEATVEQVAVKVAANVKNALGNELQKVTVDGLTALADTV----- 187
115 SMPENLMMLOIYOQDEGLISGAQANELLRTGEPVPAEYVHL-MDNHFPANDPMKMNQ 173
188 -APDTTASTLVGDGE-GTLIDSTTTATPSMAEPNGAALAPHTLRQDGAKATSEPMW-- 243
174 GYPNNKEPLLPGNNAH---YTCGVATAAQAQIMRHSMPLOGESFPDHYAGSLVGNMST 230
244 -YFESYDLXSYNNMASTYK-----AEVDATIRSLDNDSTTAVLAELVSRITGD 236
231 FGEYWDIMNPG-----NDLNLQSOVD-AVATLMRDVSAVSXSFYENGSGTYSYV 283
297 VLEKY--TIPEGESVTFSHPTKVNANNSITVYDPSLANSANPGALKEFSAN-DVYSTI 353
284 VGCALR-NMRY-----KRSLOL-----HYRALYTSQEM---HMINGELA 320
354 IVPAYQINTRYTESKVALATGLQTIAGQVTPSSVR-VFQYDVTATTAAGVGPYP 412
321 SGRRPYVAG-----NNOSIGHAFVCDGASDGFHFHMGMGV--SNGF 362
413 KG--TVYLAGVQKDYQKYRIEIVENDQ---VLKFYLDPTYKGEVDMGCTTGTG 466
363 YKLTLSPTSLIGSGIGFTTYQELITG-----IEPAKT----- 397
467 IELLTSPITTKVG-----TIYDYNINSKITAPFTIDPKVMYFKESEONEOGSKYV 520
398 -----PAAGT----- 403
521 IAGMSDETTKGIYKIYATQVWTKLGTNEMGWEDYSDDOAGIKFNKGFMPACVONT 580
404 -DALPILAK-----DIAEYKSSGSLNAGVSIYNTGEE---OSNLDLGR 445
581 LKNATPATAVETTYIKESSKYGVYDVEYDQKQIVNSVYDTPKSALETGYNVDVRR 640
446 ---LNKADGEV-----IEVKTSSINISW----- 465
641 PASLVAADTVFYFYKVKSDAKTGTCTVYAKYVTEKAGSVANVNEVDINGKVIKAP 700
466 -----YGT-----GEHP-----ES 474
701 VSEDKAPKPGVNDPDLQKLASITFEKREYKLVAPAGDPVGVKGNLIEVGNNITAKG 760
475 PSLANOLSGGIN-TITLLYRRTGTEOMEPRVHAAGGVNISKVN-----TTDP 522
761 IDPTTGKLEAGVNAKEVTVYRAVTSYVYVNYKDTGGINIKDPEVDSDAPVADATYTDK 820
533 N--NVVVTVDNN-----EGKLSI-----VPSFVADLNSYHS 553
821 KPNELITDGSRYVLPVKTDGEENGKYEIGITITTYVYQYKANKIP--IPNVEPTRP 878
554 TITVOPNSDPELTPVAFALSTGATADYISLGVNAEYVGGSSNPVYWSKVDLTLS 613

```

```

879 KVPPEPPEPDEPID-----TTPGTNGEVPNIPY-----VPG-----YTPVPKDNTPRK 925
614 EGDYTLMYRPSYNOKDEMKNIGSV-----SVKPTETHTHLEFVGNHGTSTYVLDMA 666
926 PID-----PNDPGK--GYVPTPEPNPVDPIPIYV-PVKVYVNNH-----VDEE 966
667 HNVLP--DPTLKNLGLP---ENGELVYVFPROTOSSGSLMAAQTETHIKGE----- 714
967 GNDIAPQEEGTPKPKNSIPGYEFYGTIV-----TDEGNTHTHYKKTPEVKNGTVVVNYVT 1021
715 ---TFYKRPVVEGPI-PDGS-YRATLHAFVNGQOLYLKGRNRYTKYNGTAVAEISS 769
1022 EDDTVAKEPYTDPTPSPGTPYDTT---DNKPKITFKGE-EYELVRVDTGENCKVVEG 1076
770 EEI-----RVFPNADRYV-----EISAPCIPOE-----TSILFDLSGKIYMKNSLSA 813
1077 ETVVTVYVRKVEVTPAKKVVYTNHVEDGNPNVAPQEBSTKPKNSIPGEFFGKIVTDEGNT 1136
814 GH 815
1137 TH 1138

```

RESULT 13

D85724
 hypothetical protein ydek [imported] - Escherichia coli (strain 0157:H7, substrain ED
 C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 14-Sep-2001
 C:Accession: D85724
 R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; May
 Iller, L.; Grobeck, E.J.; Davis, N.W.; Llm, A.; Dimalanta, E.; Potamousis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
 A:Reference number: AB5480; MUID:21074933; PMID:11206551
 A:Accession: D85724
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1343 <STO>
 A:Cross-references: GB:AE005174; NID:g1251519; PID:AA056256.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain 0157:H7, substrain EDL933
 A:Genetics:
 A:Gene: ydek

Query Match 3.3%; Score 144.5; DB 2; Length 1343;
 Best Local Similarity 21.7%; Pred. No. 1.3;
 Matches 155; Conservative 83; Mismatches 245; Indels 231; Gaps 40;

```

161 NGHFNPMKRNNGTPNNKEPL---PNGNHAYTCVATAAQAQIMRHSMPLOGESFD 217
295 NGYFNGAVNINSNGLINNKESLVGVQDGSHTVNVYTDK----- 334
218 YHAGSLVGNWS--GTFGE--MYDMINMGNDLNLQSOVDATATLMRDVSAVSMSFY 273
335 -----GHMNFPLGT-GEAFRTYITIGAGXGELVNSREGRVD-----SGITTGMR 377
274 ENGSGTYSV---YVGCALRNPNFRYKRSLOLHV--RALYTSQEHNDIMRGLASGRPYV 327
378 ETGNGNLTVKDKNSVITNLGTNLGDGHEMNINBGLVYNSNGSSSLGGETGKASTY 437
328 AGNNSIG-HAFVCDGASDGFHFHMGV-----WGVNSGFYKLTLLSPTS--- 372
438 TGGIWEVKNKRYTTIGVAGVGNLINSIDGKRVSONITFLDKKASIGITVLMNATSSFD 497
373 ---LGIGEGEIT-----GFTYOE--ITIGIEPAKPAEAGDAILLAKIDEAYKS 420
498 VGIVNGNFGSIVAVNSGAILNLSGYFTIGNMSGKIYVISTDSIMNLKTSSTNML- 555
421 ESSLNAGVSIYNTGEOSNLDLGYRLNKADGEVLEVTSSI-----NISWYGYGEPH 472
556 ---LVYG--VLGTGE-----LNTTGGIYKADPDQIALNDSKSGVRRVQDGNLSL 600
473 ESFSLANOLSGGIN-TITLLYRRT-----GTEQW-----EPV-----RHAOGGYV 512

```


Matches

	156:	Conservative	89:	Mismatches	271:	Indels	275:	Gaps	41:
OY	47	TAVSDKSIDVYRQDAE-----RGITSOEGSPAY----	FVYANRGNNEGVALVAAD	97					
Dd	370	TFTNMPAIVNVYVSANDYKLTSTFKDAQOGEELAPVDYTKDYHI----	QDNVTYTTNA--	422					
OY	98	RIFPIILASPIGRDMSDPNDLRLMWIQLIIDGILLSGKAQNEILTEGVPAEVHA	157						
Dd	423	--TIPGSVL-----STRPN-----QNGTGAGVTVNYV-----	450						
OY	158	LMDNGHPANDPMRRNOGPWNN---KEPLLPGNHAYTCGVAATAAQIRYH--SWPLQ	211						
Dd	451	-----KKDKDTLTSTYKDTNGOELKAPVVDATTATTHYODPTTTAAVFPGYLVAFTP	504						
OY	212	GGGFEDAHAGSIVGM-----SGTFEMTDWIMPNPDLNLTOSQVAYATLMAR	262						
Dd	505	ATGTFFG--SSNITVINYQAQGYOLSTFFDOOGKTAP--DDVAKTHVANDPYTTAA	560						
OY	263	DVASVASMSTYEENGSGTYSVYVGALNNFRYSLOLHVRLALYTISOEWHDMIRGLASG	322						
Dd	561	TIREGYTLVTPANNQGFG-----TSIDIVDYKYKAEDETLTSTYKDAQG	605						
OY	323	R----PV-----YYAGNNOSI-----GHAFVCDDGASDGTFFHMNGWGVSNGFVKTL	367						
Dd	606	KELKPVPDSKATHIODNTSTAALTPTGYTLVATPANETGETHTS---DYTVANYVKLLD	662						
OY	368	LSFTSLGIGEGI-----GFTIYOELITGIEPAKTPA-EAGTDALPIILA	410						
Dd	663	LKITLSTYKDAQOGEELAPVDSKTYHIODNYATTAAVIGPYTLVATPANOSGT-----	717						
OY	411	LNDIEAKSESGLNAGSYINTGEOGNSMLDGRLNRKDGEIEVKTSSINISWGYCE	470						
Dd	718	STDIOVNYVYA--VAKLITSTYKKNQONDLA-LPKYD-----	752						
OY	471	HPESEFLARNOLSGINTLTLLRYRTGTOMEVRHAQSGYVSIKVTNDENNVYTVD	530						
Dd	753	-----TKTY-----HIODGYTS--DIAIGYTLVAAP	778						
OY	531	NNE---GKLSTVPNSFVADLNSYEHSTITVOFNSDPSPEIRTPAFA-----LST	577						
Dd	779	TNOTGTFGASDVTVN-YVYKANDY--ILTSTYKKAQGELEKTPIIDOKYHINDITYTT	834						
OY	578	GATADVVISLGWMAVEPG-----GSSNPV-VMKCDVLTLSGDYTLMYRESINNOK	629						
Dd	835	GATIP-----GTLVAAPANOSGTFGAANVTNYVYKAD-----DYTLSTYKADANGK	882						
OY	630	D-----EMMKIGSVSKPTLEYTHLPFEYGHNOTSTY-TLDMAHRVLP--DET	675						
Dd	883	EIKAPVDSKTYHTKDNYSTSAALIPGT--LVANPANOTGFENFSANTVYVYKANERYT	940						
OY	676	I-----KNL-----GLPENGELVVYFROTOSGSSLMAAOEYV	708						
Dd	941	LSTSTKNVQGEELKAIYKGFIIKDGATSGVTIPGTLYA--TPSNKRKCTFGASNTVY	998						
OY	709	HIKOGTEFYVK	719						
Dd	999	N-----YVYK	1003						

RESULT

17 J00420
beta-1,3-glucanase A1 precursor - Bacillus circulans
C:Species: Bacillus circulans
C:Dates: 07-Sep-1990 #sequence.revision 07-Sep-1990 #text.change 15-Oct-1999
C:Accession: J00420
R.Yahata, N.; Matnabe, T.; Nakamura, Y.; Yamamoto, Y.; Kamitaya, S.; Tanaka, H.
Gene 86, 113-117 1990
Article: Structure of the gene encoding beta-1,3-glucanase A1 of Bacillus circulans W
Reference number: J00420; MUID:901B5240; PMID:2311931

RESULT 17
J00420
beta-1,3-glucanase A1 precursor - *Bacillus circulans*
C:Species: *Bacillus circulans*
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Oct-1999
C:Accession: J00420
R:Yahata, N.; Watanabe, T.; Nakamura, Y.; Yamamoto, Y.; Kamimura, S.; Tanaka, H.
Gene 86, 113-117, 1990
A:Title: Structure of the gene encoding beta-1,3-glucanase A1 of *Bacillus circulans* W
A:Reference number: J00420; MUID:90185240; PMID:23111931
A:Accession: J00420
A:Molecule type: DNA
A:Residues: 1-682 <YAH>
A:Cross-references: GB:M34503; NID:g142972; PIDN:AAA22474.1; PID:g142973

A:Experimental source: strain WL-12
C:Comment: This enzyme, together with chitinase, is crucial for hydrolyzing yeast and fly
C:Genetics:
A:Gene: glcA
F:138/Domain: signal sequence #status predicted <SIG>
F:39-68/Product: beta-1,3-glucanase A1 #status predicted <MAT>

Query Match 3.2%; Score 142; DB 2; Length 682;
Best Local Similarity 19.8%; Pred. No. 0.66; Mismatches 197; Indels 184; Gaps 32;
Matches 114; Conservative 82;

```
QY 308 SQEHMDIRKELASGRPVY-AGNNOGIGHAFVCDGASDGFHNNNG-MG-GVSNFYK 364
D 78 SAFMND-----VSDGVNKKVGNMWDIDAGGYINQNGHMSDGFNGY- 125
QY 365 LTLSPSLGIGGGIGF-----TIYOEI-----ITGIEPAKPAEAGTDALPILALDIE 415
D 126 FTLSATTEIOLYSFANGVLEOLYFONINKTTITANNPTOGP-----QIT 171
QY 416 AEYSESGLVNGYSIVYNGEHSNLDGLYRLNKADGEVIEKTSIN-ISMVYGEHPES 474
D 172 ASFTGGAGFT--YTFNNDSAYT-----YEAADDLKYVVPVNSSHIDIDNNAAS 221
QY 475 FSLAPNOLSGINTTLTYRRTGEQEPVRAAGGYNSIKVTTDPNNVYTVDNNEG 534
D 222 GWIYDHNGO-----FTDGGGYWFNTES-----INVLESKTSANLVYITFNEP 269
QY 535 KLSIVRSFYADLNSYHSHITVVOFNSDPEITPAFALSTGATADVYSLGWMAEV 594
D 270 ---TRNSY--ITPYGFTTADANG-----SIDIPIPKI 299
QY 595 PGSSNPVYVMSKDVLTLEADYTLWRFSSINNOCKMKIGSVYKTPREYTHLPREVG 654
D 300 DCGA--PI--AKEL-----GNEY--YQINNGO--W-----VDLSNDSOSKRAYSANG 339
QY 655 HNOFTST-----YTLDMAN--RYLPDFTKNGLPENG-----LVVFRQTOSS 697
D 340 YNNNSDANQMGYADYTYGLMFOPIQENMOIRIGYPLNGAGNIGNVYTFIGAPNA 399
QY 698 SCSLMAQOEYVHK-----QGETFYKPYVESPIPD-----GSY----- 731
D 400 PRPDVSQOEDISIGTPDPAIAGMNLIMODEFNCTLIDTSKNWETYYLINDPATWGW 459
QY 732 RATLHAFVNGOQOYLTK-GKRNATYKLVNGTAVEAISSEIRVFPAPADYVISA PCI 790
D 460 NAELOHTNSTONTYVDDGLN-----IKAMDSKSFDPDPKRYAQS----- 502
791 PETSITLFDLSGRIYVKNLSAGHGMV-SRLPNG 826
503 -----SKRINTKDKLSTKGRVDFRAKLPTG 528
```

RESULT 18

T20492
hypothetical protein F02C12.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T20492
R:Colles, L.
Submitted to the EMBL Data Library, September 1995
A:Reference number: 219282
A:Accession: T20492
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1827 <MIL>
A:Cross-references: EMBL:Z54265; PIDN:CAA91022.1; GSPDB:GN00028; CESP:F02C12.1
A:Experimental source: Clone F02C12
C:Genetics:
A:Gene: CESP:F02C12.1
A:Map position: X
A:Insertions: 34/1; 90/1; 113/2; 182/3; 253/3; 301/3; 385/3; 476/3; 521/3; 559/3; 598/1; 62
C:Superfamily: *Caenorhabditis elegans* hypothetical protein F02C12.1

Query Match 3.2%; Score 141.5; DB 2; Length 827;
Best Local Similarity 20.5%; Pred. No. 0.96;
Matches 127; Conservative 85; Mismatches 239; Indels 169; Gaps 30;

```
QY 334 IGHAFCVCDGASDGFHNNMGVSNNGFYKLLTLSPSLGIGGGIGFTYOEITGIE 393
D 155 IGAIVTQKYKYGIGLARLDW-----FQIDKDEEQMKNLGRKIGFAVYQ-----Y 199
QY 394 PAKTPAAGDIALP-----ILALDIAE-----YKSESLNV-----GSIYNGEEQ 437
D 200 PTNLPOEERNSLMDQLOELMKTMELKAEANNNSKSKKIDLFVEPFGT-----EQ 254
QY 438 SN-----LDGYRLNK-ADGEVIEKTSINISMVYGEHPESFSLAPNOLSGIN 487
D 255 EDHILKRLDELFDINRQKKOKESESINVEKFPVFPADGSNLIPLICLRIS---D 311
QY 488 TITLLYRRTGT--BOMEFVRAAGGYNSIKV---NTDPNNVYTVDNNEGKLSIYVNS 542
D 312 TKIMYESTGAYDWE-----DFLNKKNLPCNLIYPDGLYEV---QKIKLKENH 361
QY 543 FVADLNSYHSHITVVOFNSDPEIRTPVAFALSTGATADVYSLGWMAEVPDGSNYP 602
D 362 FVFSWNEVYVLTCAFTLGEVSFQISDRSAMGIGVAAITGIVGLFTPLAPVAATTLTYT 421
QY 603 VVMSKDVLTLESGDYTLWRFSSINNOCKD-----EMKKI-----GSVGYKT 642
D 422 AI-GTGYSMARSGHYHYDRYIHNNVNPPLSRNPDMLAVALASLAFGIGSAYL-- 478
QY 643 PTEYTHPLFEYGHNOJSTYTLDMANRYLPDFTKNGLP-----FNGELVYFR 692
D 479 -TVMTQGLEV--SOALEYTVNAA---IFANFAVSGIALTPASGNIYFEKVANGRPPL 532
QY 693 QTOSSSSLMAAQTYVHIKQGETFYKPYVNGPFPDGSYRATLAFVNGO----- 742
D 533 LFQSTSLLEFTNAVNLQAEKLIQOTQD-KIND--YRDLTKTAEDKQDFDKINEM 589
QY 743 ---QOLYLKGRNY--TVKIVNGTA-----VEAIESEEI--RVFPAPADYVISA-- 787
D 590 KNTGQOQNSKMKANLEITIDIHKAMLADKVEVHIEHVDYPRKSKPSPSSVYKMAQK 649
QY 788 -----PCIPQETSITL---FDLSGRIYVKNLSAGHGRMD 819
D 650 LTVSDRKIVEEHLQALDARGSKLPASOKITTYLVNGMDIAKV--KALLSGAGGAL 707
QY 820 VSRLPNGAYILKVDGYTTKI 839
D 708 TERL-----IDGKTKV 719
```

RESULT 19

A90428
serine protease, subtilase family, probable [imported] - *Sulfolobus solfataricus*
C:Species: *Sulfolobus solfataricus*
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: A90428
R:Shen, Q.; Singh, R.K.; Contaloni, F.; Zivanovic, Y.; Allard, G.; Ayaer, M.J.; Ch
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi Ngoc, H.P.; Redder
arrelt, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.
Submitted to GenBank, April 2001
A:Description: *Sulfolobus solfataricus* complete genome.
A:Reference number: A99139
A:Accession: A90428
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1308 <KUR>
A:Cross-references: GB:AE006641; NID:913815855; PIDN:AAK42680.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO2551
Query Match 3.2%; Score 141.5; DB 2; Length 1308;
Best Local Similarity 18.5%; Pred. No. 2;
Matches 189; Conservative 137; Mismatches 341; Indels 355; Gaps 53;

```

Query Match      32%; Score 141.5; Db 2; Length 1771;
Best Local Similarity 19.9%; Pred. No. 3.1;
Matches 103; Conservative 95; Mismatches 262; Indels 299; Gaps 41;

QY      95 ADDRIPITLAVSPIGRFDMDSPDNRLML-----QIYQEIIGLILSKAQLNEELIATE 149
      347 ADIRISTIGLVA-----DNLALSTPESQIYLORIGV----- 379
      150 GVPAEVAHMDNGHFANDPRMNQGYPMNNKEPLLPGNHVATGCVATAAQIMRHSWP 209
      380 -----DFENIP-----PLAPGOSFNIGSLAVSAGQ----- 405
      QY      210 LOGGSEFDYAGSLVGNWMSGTEGEMDTNMGANGDLDLNTQSOVDVATLML-----RDVS 265
      Db      406 -----SFEV-----DEPDWLG--GSKKGGISAVRIDAVILKIPVGDILA 444
      QY      266 ASVSM-----SEYENGSGTYSVYVVGALRNFRFRKSLDHLVHALTISOEW--HDM 314
      Db      445 AKSILIGFPVAESRFLNLDAAIKGTGV--LKNITTYIRPFEDV-----ITTNGEFSIDAR 498
      QY      315 IRGELASGR---PYV-----YAGNNOISIGAHFVCDGYASDGFTH-----FNMGSGV 358
      Db      499 LNFSLSGARLNIPPIFPLGGLQLAGNNAIKKA---APYHPTAOFRRPEYGVAGWVDV 555
      QY      359 SNQFYLLTLSTSLGIGEGIGFTIYQEIITTGIEPAKTPAEGCTALITLAKIDIEAV 418
      Db      556 NLLFEFDSI-----GFRVGFQKGVGIKYGEDIKGEGIALTYDSDLEFGL-----A 601
      QY      419 KSESGLNAGSYSTYNTGEOSSMLDGLYLNKADEVEIVTSSINISW----- 465
      Db      602 TSESAL--AAEEQNAGDQLAKQSFVNRQGR-----EMASVISRLRLRGRIENPAAAF 652
      QY      466 -----YQYGEH-----PE-----SFLAPNQL 482
      Db      653 NLKSPSGTVYRYADALQGRVYKLIMNEFSDEFGSLFIPREPAGTWTYQVNSDDIGPQI 712
      QY      483 SGGINTITL-----YKRTCEQWEVVRRAAGG--YVNSIKVN---TTDPNN-----VYVT 528
      Db      713 KMNQNSIAPVAELIKQFARDPNRQMWDLTYTGVDFPAEIKVKNLADIDGCGGGEFLITT 772
      QY      529 VDN-----EGKSIIVPNS---FVADLMSYHSHITTYQFNSDSEDEIRPPVAFALS 576
      Db      773 IDHDPKSSGSYRYRMLDSNPQSYTHLKAASAESGHLNDRRYTFGD-----LSVANS 824
      QY      577 TGAATADVISLGMVALEPVGSSNTPVYWSKDYVLLISEGDTYLLMYRSINNOQDEBKKTG 636
      Db      825 TGAHTSNLPIGTIEL---GQGNLDPVY--KDL-----SAN-----WIGG 860
      QY      637 SVSVKTPTEYTHPLEFVGHNGTSTYTLDMANNRVLDPFTLKNGLFNPSELVYVFERQTS 696
      Db      861 KFSVTSWDIATPQYYINVRKSNNPAGVLD--NPLDIPDS---GI-----Qp 900
      QY      697 SSGSLMAAOEVIHAKQGETFYVKPVEVGGPIPDQSDYSRATLHAFAVNGQOQLYLKGRNRYVK 756

```

Db	901	TSGLV-----TVEEDGRFYHRLIVDGLTEGELYRFQIQS-VDDNDRLGAVGKTSILAV-	953
Qy	757	IVNGTAVEAIESSEIRNFRNPADYVEISAPCIPQETS	795
Db	954	AGNYDAIALGESDE-----WEYVAVIGETTYTREVVS	984

Search completed: May 29, 2003, 13:27:24
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:23:06 ; Search time 15 Seconds
(without alignments)
2330.970 Million cell updates/sec

Title: US-10-030-330-1

Sequence: 1 MKKSLFLAIVMLFGIAMQGH.....PNGAVILKVDGTTKRNIVH 843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 112892

Maximum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1198	27.0	868	1 PRTT_PORGI	P43158 porphyromon
2	494	11.1	398	1 SPEB_STRPY	P00788 streptococc
3	179	4.0	939	1 EAE_ECO27	P19809 escherichia
4	154	3.5	1325	1 YDEK_ECOLI	P32051 escherichia
5	146	3.3	1409	1 HAP1_HAEIN	P44596 haemophilus
6	144.5	3.3	1256	1 MRP_STRSU	P32653 streptococc
7	142.5	3.2	1321	1 AB11_MOUSE	Q9QY93 mus muscull
8	142	3.2	682	1 E13B_BACCI	P23903 bacillus ci
9	139.5	3.1	814	1 KEX2_YEAST	P13134 saccharomyc
10	139.5	3.1	890	1 BCNS_CLOPE	P08686 clostridium
11	138	3.1	1398	1 PLS_PYRPU	P72186 pyrococcus
12	136.5	3.1	3354	1 CADN_HUMAN	Q9H251 homo sapien
13	136	3.1	426	1 GUN2_ERWCH	P07103 erwina chr
14	136	3.1	490	1 TAY3_TREDE	P18164 treponema d
15	135	3.0	987	1 YD91_METJA	Q58789 methanococ
16	134.5	3.0	976	1 AMY_BUTEI	P30269 butyrivibri
17	133	3.0	837	1 HFC1_HAEIN	P33397 haemophilus
18	133	3.0	934	1 EAE_ECO57	P43261 escherichia
19	132.5	3.0	802	1 CSD2_ECOLI	P53513 escherichia
20	132.5	3.0	1103	1 VQ37_BPARI	Q09055 bacterioph
21	130.5	2.9	1286	1 AIDA_ECOLI	Q03135 escherichia
22	130	2.9	762	1 SLAP_ACEKI	P22258 acetogenium
23	130	2.9	914	1 GUN2_CLOSR	P50990 clostridium
24	127.5	2.9	3317	1 CADN_PAT	P58365 ratius norv
25	125	2.8	837	1 HFC3_HAEIN	P45998 haemophilus
26	124.5	2.8	878	1 FIMD_ECOLI	P30110 escherichia
27	124.5	2.8	820	1 CHIA_ALTSO	P32883 alteromonas
28	124.5	2.8	865	1 HTRE_ECOLI	P33129 escherichia
29	124	2.8	1541	1 IGAL_HAEIN	P42782 haemophilus
30	123.5	2.8	852	1 CSG_HALNA	P08198 halobacteri
31	123.5	2.8	862	1 SLA2_BACAN	P94217 bacillus an
32	123	2.8	676	1 HMUR_YERPE	Q56969 yersteinia pe
33	123	2.8	702	1 FOXA_SALTY	Q56145 salmonella

34	123	2.8	935	1 EAE_ECOLI	O31000 escherichia
35	123	2.8	1259	1 YTFN_ECOLI	P39321 escherichia
36	122.5	2.8	1218	1 MGPC_MYCPN	O50341 mycoplasma
37	122.5	2.8	1609	1 FIG2_YEAST	P25653 saccharomyc
38	122	2.7	760	1 YBIL_ECOLI	P75780 streptococc
39	122	2.7	936	1 EAE_CITER	O07591 citrobacter
40	121.5	2.7	1628	1 NAGH_CLOPE	P26831 clostridium
41	121	2.7	837	1 HFC2_HAEIN	P45997 haemophilus
42	121	2.7	1167	1 CLAA_BACTU	P56956 bacillus th
43	120.5	2.7	819	1 CSD1_ECOLI	P53512 escherichia
44	120.5	2.7	972	1 CTAL_BACCI	P94286 bacillus ci
45	120	2.7	1953	1 BIGA_SALTY	P25927 salmonella

ALIGNMENTS

RESULT 1
PRTT_PORGI STANDARD; PRT; 868 AA.
AC P43158;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Thiol protease/hemagglutinin prt precursor (EC 3.4.22.-).
GN PRTT.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides; Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 53977;
RX MEDLINE-93114862; Pubmed-8093357;
RA Otago J.-I., Kuramitsu H.K.;
RT "Isolation and characterization of the Porphyromonas gingivalis prt gene, coding for protease activity."
RL Infect. Immun. 61:117-123(1993).
RN [1]
RP REVISIONS, SEQUENCE FROM N.A.
RC STRAIN-ATCC 53977;
RX MEDLINE-95105001; Pubmed-7806362;
RA Madden T.E., Clark V.L., Kuramitsu H.K.;
RT "Revised sequence of the Porphyromonas gingivalis prt cysteine protease/hemagglutinin gene: homology with streptococcal pyrogenic exotoxin B/streptococcal proteinase."
RL Infect. Immun. 63:238-247(1995).
CC -1- FUNCTION: APPEARS TO BE SPECIFIC FOR ARGININE-CONTAINING PEPTIDE BONDS. POSSESSES HEMAGGLUTININ ACTIVITY.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C10.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use, by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: M83096; -; NOT_ANNOTATED_CDS.
DR MEROPS: C10.002; -
DR InterPro: IPR000200; Peptidase_C10.
DR InterPro: IPR001230; Prenyl_site.
DR Pfam: PF01640; Peptidase_C10; 1.
DR PRINTS: PRO0797; SYRPEPTAPIN.
KW Hydrolyase; Thiol protease; Signal; Hemagglutinin.
FT SIGNAL 1 27
FT PROPEP 28 ?
FT CHAIN ? 868
FT ACT_SITE 184 868
FT ACT_SITE 327 327
FT ACT_SITE 327 327
SQ SEQUENCE 868 AA; 96444 MW; 45436FEF32779323 CRC64;

Query Match 27.0%: Score 1198; DB 1; Length 868;
 Best Local Similarity 33.9%; Pred. No. 2,2e-65;
 Matches 290; Conservative 155; Mismatches 340; Indels 70; Gaps 23;

OY 1 MKSFLAIYMLFGIAOGHSPATYKERALSAR--LALQVSLRMGOTAVSDKISIDYV 58
 DB 1 MKRIFLYLGLLLCLPLM-LQAGPTRSKAEOTAKNFAKQPTL---SSSTASLRMDFV 55
 OY 59 YRGDDAREGITSQEESSPAYFYVARNNGNEGVALLVADDDIPRTILAYSPIGRDPMQMD 118
 DB 56 YKAAREE-----ALFEVFNKEKDGFLVADDPPEYIGIAFGHEDPAARMPD 105
 OY 119 NLRMWLOIYDOEIGLILSGRAQLEELR---TEGPAEYHALMDNGHFANDPMRNQY 175
 DB 106 NLGMLKGYREMLAVMDKAEPIDPIREAKPTRDLPSSIAPILETGEHMSDPLTMQGY 165
 OY 176 PMNKEPRLPNHAYTCVATAAQIMRYHSPLOGESSFDYHAGSLVG--NWSGTFG 232
 DB 166 PFYTLHLPLPSGOAAYTCVATAMGOIMRYHKPEKSGCEYDYV-DMTGTHTHYSGTFG 224
 OY 233 EMTDWMIPGNPDLDNLTOGODAVATLMDRVSASNSPYENGSGTYSYVYGALNNF 292
 DB 225 ETTNMSKMPGNISY-GISPEVKALSTFMKDVSEVNMOPADGSGFESIFVERALRETF 283
 OY 293 RYRSLOLHVALYTSQEMHDMIRGELASGRPYTYAGNNGSIGHAFCDCYASDGTFFHN 352
 DB 284 HYKSLRYIHRSLPLGKEMKDMIRKELAEKNRPYYAGADSGMGAFCDCYEPDGTFFHN 343
 OY 353 WGWGNGVNGFYKLTLSPLSGLGEGIGTIIYOEITIGTEPAKTPAEATDPLIALK 412
 DB 344 WGWGMSNGNFYLLNLPLPSGLGAGDGGYSTQEEVYIGLEPASNE-PGIVPPTLYL 402
 OY 413 DIAEYSESEGLWNGYIYVTGEEQSLDGLYRLNKADGVEIEKTSINISWYG-XGEH 471
 DB 403 GLQ-HNMSDALDLQSVIKNYSTYAGVKLALRLTLNGETETNPATVITVIMEDIGES 461
 OY 472 PESFSLAPNOLSGINTTILYRTGTGEOMEPVRAHAGGYVNSIKVTTDPNNVYTVDN 531
 DB 462 TGNITTPCSOFAEKNITISILYRTDGMADKELKHIIMGLVNNKIEV-TMPAGDVAYSV- 518
 OY 532 NECKLSIVNPFVADLNSYEHSTIYVQFNSDSDPEITPAFALSTGATADVYSLGWV 591
 DB 519 ADARIVYKDGSLSHNLKAYSDDCKLSATVVPNGTEEFSSRYTFALRN--TGRLYFLGRHL 576
 OY 592 AEV-PGSSSNYPVWMSKDVLTLSGCDYTLWY--FSINNOCKDEKKIGSVKTPREYTH 648
 DB 577 VELHPGDEGEKYSLTITGLKARAGOVMLCTGDMELMDASWIELASIEVAHEHSTHS 636
 OY 649 PLFEVGHNOTSTYTLDMANRRVLPDFTLKNL-GLPENGELVVVFPROTSSGSLMAAQT 707
 DB 637 SLIVASNPQIDLLTVHRANPELPTFSITNEGATFGSKIEIVAKFSET-FFQAKEEH 695
 OY 708 VHIKOGTFYKKVNVBEP-----IPDGSYRATIAHVNQOOLY-----LKGRNNT 754
 DB 696 MSIAOGTKVLSPELANSLSYNAELFPGIITYIVR-----EQGWDIDIDFGDYR 750
 OY 755 VKIVNGTAVEAIESE--EIRVFPNPARDVEISAPCIPOETSIIIFDLGKIYMKSL- 811
 DB 751 IRLITDLSXKNGKDVSTIVLPNPAHDYVHAIPPTVAGSTLRLFDIGRQSLSTKIR 810
 OY 812 -----SAGH 815
 DB 811 ICRYASRRRTSSEGH 825

RESULT 2
 SPEB_STRPY STANDARD; PRT; 398 AA.
 AC P00788; P26296; Q54961; Q54962; Q54963; Q54964; Q54965;
 AC Q54966; Q54967; Q54968; Q57024; Q57082; Q57202; Q57211; Q57212;
 AC Q58680;
 DT 21-JUL-1986 (Rel. 01, Created)

DT 01-MAY-1992 (Rel. 22, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Streptococcal precursor (EC 3.4.22.10) (Streptococcal cysteine
 DE proteinase) (Streptococcus peptidase A) (SPP) (Exotoxin type B) (SPE
 DE B).
 GN SPEB OR SPY2039 OR SPYM18-2099.
 OS Streptococcus pyogenes, and
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID-1314, 186103;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32 AND 146-162.
 RC STRAIN-86-858, and NY-5;
 RX MEDLINE=90330563; PubMed=2198264;
 RA Hauser A.R., Schlievert P.M.;
 RT "Nucleotide sequence of the streptococcal pyrogenic exotoxin type B
 RT gene and relationship between the toxin and the streptococcal
 RT proteinase precursor";
 RL J. Bacteriol. 172:4536-4542(1990).
 RN [2]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN-Various MGAS strains;
 RX MEDLINE=94285789; PubMed=7516997;
 RA Kapur V., Topouzis S., Majesky M.W., Li L.L., Hamrick M.R.,
 RA Hamill R.J., Patil J.M., Musser J.M.;
 RT "A conserved Streptococcus pyogenes extracellular cysteine protease
 RT cleaves human fibronectin and degrades vitronectin";
 RL Microb. Pathog. 15:327-346(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A-20 / Serotype M1,T1;
 RA Wu J.-J.;
 RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SV / Serotype M23;
 RA Hong K.;
 RT "A novel cloning method used arbitrarily primed PCR";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Pericelli J.C., McShane W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C.J., Szatse S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najjar F.Z., Ren O., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus
 RT pyogenes";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MGAS8232 / Serotype M18;
 RX MEDLINE=21927593; PubMed=11917108;
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
 RA Perkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Yeasty L.G., Musser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 RT group A Streptococcus strains associated with acute rheumatic fever
 RT outbreaks";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
 RN [7]
 RP PRELIMINARY SEQUENCE OF 28-86 AND 121-398.
 RA Yonaha K., Elliott S.D., Liu T.-Y.;
 RT "Primary structure of zymogen of streptococcal proteinase";
 RL J. Protein Chem. 1:317-334(1982).
 RN [8]
 RP PRELIMINARY SEQUENCE OF 146-398.
 RX MEDLINE=76190087; PubMed=1270417;
 RA Tai J.Y., Kortt A.A., Liu T.-Y., Elliott S.D.;
 RT "Primary structure of streptococcal proteinase. III. Isolation of

RT cyanogen bromide peptides: complete covalent structure of the
RT polypeptide chain.";
RT J. Biol. Chem. 251:1955-1959(1976).
RN [9]
RP FUNCTION.
RC STRAIN-N2131 / Serotype M49,T14;
RX MEDLINE=99081733; PubMed=9864206;
RA Kuo C.-F., Wu J.-J., Tsai P.-J., Kao F.-J., Lei H.-Y., Lin M.T.,
RA Lin Y.-S.;
RT "Streptococcal pyrogenic exotoxin B induces apoptosis and reduces
RT phagocytic activity in U937 cells.";
RL Infect. Immun. 67:126-130(1999).
RN [10]
RP FUNCTION.
RC STRAIN-A-20 / Serotype M1,T1;
RX MEDLINE=99368817; PubMed=10456871;
RA Tsai P.-J., Lin Y.-S., Kuo C.-F., Lei H.-Y., Wu J.-J.;
RT "Group A Streptococcus induces apoptosis in human epithelial cells.";
RN Infect. Immun. 67:4334-4339(1999).
CC -1- FUNCTION: IMPORTANT STREPTOCOCCAL VIRULENCE FACTOR WHICH CLEAVES
CC HUMAN FIBRONECTIN AND DEGRADATES VITRONECTIN. ALSO CLEAVES HUMAN
CC IL1B PRECURSOR TO FORM BIOLOGICALLY ACTIVE IL1B. CAN INDUCE
CC APOPTOSIS IN HUMAN MONOCYTES AND EPITHELIAL CELLS IN VITRO, AND
CC REDUCES PHAGOCYTIC ACTIVITY IN MONOCYTIC CELLS. THUS, MAY PLAY A
CC ROLE IN BACTERIAL COLONIZATION, INVASION, AND INHIBITION OF WOUND
CC HEALING.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage with hydrophobic
CC residues at P2, P1 and P1'.CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DISEASE: THE STREPTOCOCCAL PYROGENIC TOXINS A, B, AND C ARE
CC THE CAUSATIVE AGENTS OF THE SYMPTOMS ASSOCIATED WITH SCARLET
CC FEVER. HAVE BEEN ASSOCIATED WITH STREPTOCOCCAL TOXIC SHOCK-LIKE
CC DISEASE AND MAY PLAY A ROLE IN THE EARLY EVENTS OF RHEUMATIC
CC FEVER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M86905; AAA26978.1; -
DR EMBL: L26125; AAA26979.1; -
DR EMBL: L26126; AAA26980.1; -
DR EMBL: L26127; AAA26981.1; -
DR EMBL: L26128; AAA26982.1; -
DR EMBL: L26130; AAA26986.1; -
DR EMBL: L26131; AAA26987.1; -
DR EMBL: L26132; AAA26988.1; -
DR EMBL: L26133; AAA26989.1; -
DR EMBL: L26134; AAA27000.1; -
DR EMBL: L26135; AAA27001.1; -
DR EMBL: L26136; AAA27002.1; -
DR EMBL: L26137; AAA27003.1; -
DR EMBL: L26138; AAA27004.1; -
DR EMBL: L26139; AAA27005.1; -
DR EMBL: L26140; AAA27006.1; -
DR EMBL: L26141; AAA27007.1; -
DR EMBL: L26142; AAA27008.1; -
DR EMBL: L26143; AAA27009.1; -
DR EMBL: L26144; AAA27010.1; -
DR EMBL: L26145; AAA27011.1; -
DR EMBL: L26146; AAA27012.1; -
DR EMBL: L26147; AAA27013.1; -
DR EMBL: L26148; AAA27014.1; -
DR EMBL: L26149; AAA27015.1; -
DR EMBL: L26150; AAA27016.1; -
DR EMBL: L26151; AAA26980.1; -
DR EMBL: L26152; AAA26981.1; -
DR EMBL: L26153; AAA26982.1; -

DR EMBL: L26154; AAA26983.1; -
DR EMBL: L26155; AAA26984.1; -
DR EMBL: L26156; AAA26985.1; -
DR EMBL: L26157; AAA26986.1; -
DR EMBL: L26158; AAA26987.1; -
DR EMBL: L26159; AAA26988.1; -
DR EMBL: L26160; AAA26989.1; -
DR EMBL: L26161; AAA26990.1; -
DR EMBL: L26162; AAA26991.1; -
DR EMBL: AF104940; AAD17930.1; -
DR EMBL: AB030578; BAB16027.1; -
DR EMBL: AE006625; AAK34706.1; -
DR EMBL: AE010112; AAL8559.1; -
DR EMBL: L26129; AAA26995.1; -
DR PIR: A37768; A37768.
DR PIR: A00978; BPSOP.
DR MEROPS: C10.001; -
DR InterPro: IPR000200; Peptidase_C10.
DR Pfam: PF01640; Peptidase_C10; 1.
DR PRINTS: PR00797; STREPTOPAIN.
KV Hydrolyase; Thiol protease; zymogen; Toxin; Signal; Complete proteome.
FT SIGNAL 1 27
FT PROPEP 28 145
FT CHAIN 146 398
FT ACT_SITE 192 192
FT ACT_SITE 340 340
FT VARIANT 8 8
FT VARIANT 17 17
FT VARIANT 80 80
FT VARIANT 111 111
FT VARIANT 137 137
FT VARIANT 154 154
FT VARIANT 193 193
FT VARIANT 211 211
FT VARIANT 293 293
FT VARIANT 305 305
FT VARIANT 308 308
FT VARIANT 317 317
FT VARIANT 384 384
FT VARIANT 394 394
FT CONFLICT 84 85
FT CONFLICT 169 169
FT CONFLICT 187 191
Query Match 11.1%; Score 494; DB 1; Length 398;
Best Local Similarity 30.6%; Pred. No. 5e-23;
Matches 129; Conservative 73; Mismatches 155; Indels 64; Gaps 13;
OY 6 LIAIYMLFGIMOG-----HSAPYTKRALSLARLARLYSLRGQAVSPKISIDYV 59
DB 10 LLSIALOGFYLANVFDONFARFKRAKSAITFIQSAIKAGASAED-ITLDKVN 68
OY 60 RQGAERNGITSQESPAFYVANGNNEGVALVAADRIPITLAYSPIGRDMSMDN 119
DB 69 LGGEL-----SGSNMYVYNISG--GFVIYSGDKRSPILLIGTSGSPDANG-KEN 116
OY 120 LRMLQIT-----DQELGLISGAQUNELRLREGYPAEYHALMDNGHRANDPMRNO 173
DB 117 IASFESYVEQIKENKRLDTTYAGTAETKOP-----VKSILD-----SKGIHNO 162
OY 174 GYPNNKEPLLPN-----GNHAYGCVATAAQAQIMRHSNPLQEGSGFDY----- 218
DB 163 GNPIYLLTPVLEKXKPGQSOFVGGHAATGCAVATATQIMKHINPKLKYTYTLLSSNN 222
OY 219 ---HAGSLVGNWSGTGEWTDWIN-MGDNPDLDNLTSQVDAYATTLNRDVSASVNSFY 273
DB 223 PYFNHPKML---FAAISTROYNMNIIPTYSGRRESNVQKM--AISELMADVGISVDMDYG 277

```

Oy 274 5NGSGT5VYVYVGLRRNERFKRSJQLHVRALYLSOEWHMDIRBLASGRVYYAGNNS 333
Db 278 PSSGAGSSRRORALKENFNKGNOSHOJINRDSFSDMEADIDKELSONDPYVYQGVK 337
Oy 334 1GAHFVCDGYVSDGTFHFHNMCGVSNQVFLYTLTLLSPYSLGIGBGEIGFTTYOELLTGE 393
Db 338 1GHAFFVIDGADGRNPFYHNMGMGVSDGFPLDALNPALCTGGGAGAGFNQSYAVGK 397
Oy 394 P 394
Db 398 P 398

RESULT 3
EAE_ECO27 STANDARD: PRT; 939 AA.
AC P19809;
01-FEB-1991 (Rel. 17, Created)
01-FEB-1991 (Rel. 17, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
InflmIn (attaching and effacing protein) (Eae protein).
EAE OR EAEA.
Gn Escherichia coli O127:H6.
Os Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Cc Escherichia.
Ox NCBI_TaxID=168807;
Rn [1]
Rp SEQUENCE FROM N.A.
Rc STRAIN=O127:H6 / E2348/69;
Rx MEDLINE=98254123; PubMed=9593291;
Ra Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
Lai L.C., McManus B.P., Donnenberg M.S., Kaper J.B.;
"the complete sequence of the locus of enterocyte effacement (LEE)
from enteropathogenic Escherichia coli E2348/69.";
Rt Mol. Microbiol. 28:1-4(1998).
Rl -1- FUNCTION: NECESSARY FOR THE PRODUCTION OF ATTACHING AND EFFACING
CC LESIONS ON TISSUE CULTURE CELLS. BELIEVED TO MEDIATE ADHERENCE.
CC -1- SUBCELLULAR LOCATION: OUTER SURFACE.
CC -1- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 LYSEN REPEAT.
-----
Cc This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc between the Swiss Institute of Bioinformatics and the EMBL outstation --
Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (see http://www.isb-sib.ch/announce/
Cc or send an email to license@lsb-sib.ch).
Cc
Cc
Cc EMBL; M58154; AAA62775.1; -
Cc EMBL; AF022236; AAC38392.1; -
Cc DR InterPro: IPR003344; Big_1.
Cc DR InterPro: IPR003343; Big_2.
Cc DR InterPro: IPR003535; IntImIn.
Cc DR InterPro: IPR001304; LectIn_C.
Cc DR InterPro: IPR002482; LysM.
Cc DR Pfam; PF00059; LectIn_C; 1.
Cc DR Pfam; PF01476; LysM; 1.
Cc DR Pfam; PF02368; Big_2; 1.
Cc DR Pfam; PF02369; Big_1; 2.
Cc DR PRINTS; PR01369; INTIMIN.
Cc SMART; SM00257; LysM; 1.
Cc Outer membrane; Virulence.
Cc

```

FT	REPEAT	65	113	1YSM.	
SEQ	SEQUENCE	939 AA:	102410 MM:	783C53EB0322DEAD CRC64;	
QY	Query Match	4.0%;	Score 179;	DB 1; Length 939;	
	Best Local Similarity	20.3%;	Pred. No. 0.0026;		
	Matches 179;	Conservative 119;	Mismatches 330;	Indels 254; Gaps 48	
QY	9	IYMLEGIMOGHSAVTR-----	ERALSIALALEQVSLRMQOTVPSDKISID	56	
DB	132	LYAAGGVA--GHTNKLTKMSPDVTKSNMTDKALNYA-----	AQAASISLSQASISLND	185	
QY	57	YVYRGADERGITTSOEGSPAFVYANG-----	NEGVALVAADRIP-----	101	
DB	186	YA---KDFALGAGNQASSQLQAMLOHGTAEVNLQSGNNEGSSL--	DELLPYDSEK	239	
QY	102	ILASPIGRFMDS-----	MPDNLRMLOIYDQIG--	137	
DB	240	MLAEQVARYIDSFTYANLQAGQRFPENNLGNVFIIDQFSGDNTRLGIGGEYRDY		299	
QY	138	KAOLNEILRLEG-----	VPAEVHALMDNGHFANDP-----	MKNQGY PWN	178
DB	300	FKSSVN-GYFRMGSHESYNKKDYDERPANGDFIDFNGCLPSPALGAKLMEYQYGDVY		358	
QY	179	---KNEPLPNQNAHYTCGVAFAAQIMRHYHWPRLQSGSPDYHAGSLVGN-----	WGST	230	
DB	359	ALFNSDKLQSNQATVVG-----	VNYPIPLVTWG-IDYRGT--	GENEDILYSKQ	406
QY	231	FGEWMD--WINMGPNDLDNLQSOVDAYATLMDRVSASVSMFYENGSS-----	GTY	280	
DB	407	FRYGDKPM-----	SQLEPQVNVNLFRLTSGRYDVLQORNNHIILEYKKDHL	454	
QY	281	SVYVVGALARNFRYKRSQLNHRALYTSQE--	WHD-MIRGELASGRPYVYACNNOSIGHA	337	
DB	455	SLNTPHDINGTERSTOKIQLIVKSYGGLDRIVMDSALRSQ--	GGQIQHSSQSQAODIQ	511	
QY	338	FVCDGYASDGTFFHFNMGMGVNSNGCYKLTLLSPISLGIGEGI-----	GFTIYOEI	388	
DB	512	ALLPAYQGG-----	SNYKTKARAYDNGNSNNVLLTITVLSNGOYVDQYG	559	
QY	389	ITGIRPAKTPAA-CTDALPLIALKDIEAEYKESGGLNVGSIYVNTGEGSULDGRLN		447	
DB	560	YTFDTADTKSAKADTEAITTYA-----	TVKKNQVAQAANVPSE--	599	
QY	448	KADGEVIEVKTSSINISWYG-----	YGEHPES--	FSIAPNOLSGOINTITLY--	493
DB	600	IYSGIAV--LSANSANTNGSKATYTLKSDPKQOYVVSAKTAEMTSALNANAVIFPDQTA		658	
QY	494	--RRGTSEOMEVPRHAOGGYNSIKVNTDP--	NNVYVTVDNNEKSLSTVPSFYADLNS	549	
DB	659	SITEIKADKTAIVANGQDAITFYVYVMGDKRPVSNQOEYFTTTLGKLS--	NSTERKIDTNG	716	
QY	550	YEHSTIYQFN-----	SDSPDEIRP--	VAFALSTGATADVDYISLGMVAEVPGGSS	599
DB	717	YAKVTLSTTPGKSLVARSVDVAADVAPEVEF--	FTLLTIDD--	GNIIYGTGVKG	770
QY	600	NYPVWMSKDVLLISGDDTYLWRFESINNOCKDEMKI-----	GSVSKTPELEYH	648	
DB	771	KLPITW-----	LQYGOYNL--	KASGGNGKTYMRBSANPRLASVDASSGGVYTLKEGTTII	822
QY	649	PLEVEGHNOTSTYTLDMAHNRVLPDF-----	TLKNG--	LPRN-GEIYVYFR--	692
DB	823	SVIS--SDNOTAYITATPNSLIVPMKSRVYNDAVNCKNGKGLPSSONBLENYFKAM			881
QY	693	-----QTOSSGSL--	NAAOETVHKKOGEFVYVPRVYEGPI	726	
DB	882	GAANKYEYKSSQTIISWQDTAODAKSGVASTIDLAKONPL		923	

D	b	256	IYDAKKKKMLINAVLQGTGHPPEGRNGQLIREMFYENVALVDTPSPQRIIPPLNGHY	315
Q	y	352	NMGWGSN--GEFKLLTLLSPSLGIGEGIGFTIYOEITIGIEPAKTPAEAGTALPIL	409
D	b	316	SE----VSNNDTGKTLTLRPSKDG-----SKAKSEVGYKLFNP	351
Q	y	410	ALKDIEAFY-KSESLAV-----GYSTYNTGEQSNLDGRLKADGEV-----IEY	456
D	b	352	SLNQTAKSHVAAAGYNTYOPMEYGNKIYLDQCKGLTLENNINOGAGLYFEGNFVY	411
Q	y	457	KTSINISIVGCG-----EHDESPSLAPNLSQ-GINTI-----	489
D	b	412	KGKQNNITWQAGVISGDADATYEMKIVNHNEN-----DRLSKIGIGTLVNGKMLGSIS	466
Q	y	490	---TLVRYRTGTEMEPEVRHAGGYVN---SIKVNNT---DPNNVVYVDNNEKGLSTIV	539
D	b	467	AGNGKVLIDQDADEAGQKQAFKEVGIVSGRAIVQLNSIDQVDPNNIYGEF--RGRL---	521
Q	y	540	PNSEVADLNSEHSITTYQFNSDSDEL---RTVYARALSGATADYISLGWYAEVP	595
D	b	522	---DLNGHSLTFERRIO-NTDEGAMLYVNHNTQVANITTIIGNES-----IYAP	564
Q	y	596	GGSSNY-PVWYSKDVLTLESDDYTLMTYRFSINNOCKENKIGISVAV--KTPETYNHLP	652
D	b	565	SNKNINKLDQSKELA-----YNQW-----GEFDKKKHNKRLMLIKPTTEDRTLLS	613
Q	y	653	VGHNOTSTYT-----LD-----MAHNRVLDPF-----TLKNIG	680
D	b	614	GGTINLKGDIQTCKGLFEFSGRPTPHAYNHLDRKRWSEMEGIPQGEIWMYDWINRFFKAEN	673
Q	y	681	LPFNGELVYVRQROSSSGSLMAAQETHHINQGEFFVKKPYVEGPIPDGSGYATLHAPN	740
D	b	674	FQIKGSAVYVSRNVSISIEGN-W---TVSNNAATF-----GVVP-----N	709
Q	y	741	GOQOOLYLKGRKN--YTVKIVNGTAEVAIESESEIYVFPNPADYVEISAPICPETSITL	798
D	b	710	XQNTICRSDMTGLTCTKTVMLTDIKVI-----NSLPI	742
Q	y	799	FDLSGKIVKNSLSAG-HGRMDSRLPNCAVILKVDGYTTKIN	840
D	b	743	TQINGSINLTJNNATVNIHG-----LAKINGVTLILD	773

RESULT 6

MRP_STRSU

MRP_STRSU

STANDARD;

PRT; 1256 AA.

AC

P32653;

DT

01-OCT-1993 (Rel. 27, Created)

DT

01-OCT-1993 (Rel. 27, Last sequence update)

DT

15-JUN-2002 (Rel. 41, Last annotation update)

DE

Muramidase-released protein precursor (136 kDa surface protein).

GN

MRP.

OS

Streptococcus suis.

OC

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC

Streptococcus.

OX

NCBI_taxonomy:1307;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN-serotype 2 / D282;

RX

MEDLINE-92267650; PubMed-1587602;

RA

Smith H.E., Vecht V., Gielkens A.L., Smits M.A.;

RT

"Cloning and nucleotide sequence of the gene encoding the

RT

136-kilodalton surface protein (muramidase-released protein) of

RT

Streptococcus suis type 2.";

RL

Infect. Immun. 60:2361-2367(1992).

CC

-1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by

CC

an amide bond (potential).

CC

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC

between the Swiss Institute of Bioinformatics and the EMBL outstation

CC

the European Bioinformatics Institute. There are no restrictions on its

CC

use by non-profit institutions as long as its content is in no way

CC

modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

 DR EMBL: X64450; CAA45781.1; -
 DR PIR: A43829; A43829.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR TIGRfams: TIGR01167; LPXTG_anchor; 1.
 DR TIGRfams: TIGR01168; YSIRK_signal; 1.
 DR PROSITE: PSS0847; GRAM_POS_ANCHORING; 1.
 DR Cell wall; Peptidoglycan_anchor; Repeat; Signal.
 FT SIGNAL 1
 FT CHAIN 48 1226 POTENTIAL.
 FT PROPEP 1227 1256 MURAMIDASE-RELEASED PROTEIN.
 FT REPEAT 663 681 REMOVED BY SORTASE (POTENTIAL).
 FT REPEAT 839 861 SMALL.
 FT REPEAT 953 1006 SMALL.
 FT REPEAT 1064 1084 LARGE.
 FT REPEAT 1089 1142 LARGE.
 FT REPEAT 1143 1195 LARGE.
 FT REPEAT 1223 1227 LARGE.
 FT SITE 1226 1226 LPXTG SORTING SIGNAL (POTENTIAL).
 MOD_RES 1226 1226 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 SEQUENCE 1256 AA; 135795 MW; DCF765242F14341 CRC64;

 Query Match 3.3%; Score 144.5; DB 1; Length 1256;
 Best Local Similarity 18.7%; Pred. No. 0.49;
 Matches 214; Conservative 121; Mismatches 374; Indels 433; Gaps 60;

 QY 21 SAEVTERALSLRLALROVSLR-----MGQAVSDKISIDVYRQDAERGITS---- 70
 DB 83 SAVALTSTEVEKAKAVLEQVTSPLLAGLQKELAKTEDATLAKAIEDAQTKLAAKAI 142
 QY 71 -----OEGSPAFYVYANR--GNN-BGYA---LVAAADRIPITLAVSPGRDMD 114
 DB 143 LADSEATVEQVEQAVKAVKANEALNELKTYVQGLTALADTV----- 187
 QY 115 SMEDNLRMLQIYDOEIGLILSGKAQLENEILRTGVPAYVHAL-MDNQHFANDPMRWQ 173
 DB 188 -ADDTASTLKVDGE-GTLDSTTATPMAEPNGAIAPIHLRQDGKATSEPNW-- 243
 QY 174 GYPMNKEPLLPNGNHA---YTCGVATAAQIMRYHSMPLQEGSEPDYHAGSLVGNWGT 230
 DB 244 -YFEESDLYSKNNMASTYKQ-----AEVDAYIYSDNDSTTAVLAELVSRRTGD 296
 QY 231 FGEMYDWIMNG-----NPDLDNLQSOYD-AVATLMRDVSAVSMSFENGSGTYSY 283
 DB 297 VLEKY--TIEPGESVTFSHPTKVNANNSITVTYDLSLASANPGALKESAN-DDVYSTI 353
 QY 284 VQGLR-NMERY-----KRSIQD-----HYRALYTSQEW---HDMIRGELA 320
 DB 354 IVPAYQINTRYVYESGKVLATYGIOTIAGOVVTPSSVR-VFTGYDVAATTAKAVGPPY 412
 QY 321 SGRPVYAG-----NOSIGHAVCYDGYASDQTFHNMNGGV-SNGF 362
 DB 413 KG-TYVLAGVQKDVQYKVIKREIVENDQA---VLKFFYLDPTYKGEVDMRGDTDTGF 466
 QY 363 YKTLSPSTSLIGEGIGFTYQELITG-----LEPAKT----- 397
 DB 467 TELTTSPTYYKVG-----TIYDYNINSKITAPETIDPTKVNWKESEONEGSKYRV 520
 QY 398 -----PAEAGT----- 403
 DB 521 IAQMSGETTKGIYKIYINTCYWTKLGTNEMGMPYSDQACIKENKNGFPAQVQNT 580
 QY 404 --DALPLALK-----DIEEYKSEGLANGVSYITNGE---QSNLDIGYR 445
 DB 581 LRNATPATAVETTYIKESKSGDVYIYEDTGKQIVNSVDPKPSALCTEYNTDVR 640
 QY 446 ---LNKADGEV-----IEKTSINSIW----- 465
 DB 641 PASLVAAADGVITYYKESKSDSAKTGTGTVAGTTVYKVKAGSVNNEVDINGKVIKAP 700
 QY 466 -----YGY-----GEHP-----ES 474

DB 701 VSDEKAKKPGYKNTDOLDOKLASTFEGREKYLVPAGDIYGVKGNLIEVGNNTRAKG 760
 QY 475 FSLAPNOLSGIN-TITLLRRFTGEQMEVRRHAGGYVNSIKVN-----TDP 522
 DB 761 IDPTFKIEAGVKNVEYVYRAVTVGVVNVKDEGNVIKDPEIDVSDAPVDAVYTTDK 820
 QY 523 N-NVYVAVDNN-----EGKLSI-----VNSVADLNSIEHS 553
 DB 821 KPNELITKDSRRVYLVPSKTDGEENGKVIETITVYVYOKVANMPE--IPVPEPDRP 878
 QY 554 TIVQFNSDPSDEIRTPVAFALSTGATADYISLGVMYAEVPGSSNYPYWMSKDVLTLS 613
 DB 879 KVPFPDPTPEDEIDP-----TTPGNGEVPNIPY---VPG---YTVVDPDRDNP 925
 QY 614 EGDYTLMYRFSINNOKDEMKIGSV-----SVKPTETHTPLFEVGHNOTSTYTLDMA 666
 DB 926 PID-----PNDPGK--GYVPTPENPGVDPIPVY-PVKRVVTNH-----VDEE 966
 QY 667 HNVLP--DFTLKNLGLP---FNGELVYVERQTOSSGSLMAOETVHIQGE----- 714
 DB 967 GNPLAPQEGTKPKRSIPGEYFGKVY-----TDEDGNTTHIKRTPPEVNGVYVYVT 1021
 QY 715 ---TFVYKPVYEGPI-PDGS-YRATLHAFVNGOOLYKGRNTYKIVNGTAVEAIESS 769
 DB 1022 EDGVKIEPTDPTDPTSPGCTPIDPT-----DNKRTIFKGE-EVELYRVQDGTENGKVV 1076
 QY 770 EEI-----RVFPNPARDYV-----EISAPCIPOE-----TSLIFLDSKRYKNSLSA 813
 DB 1077 EFTVYTVYRRVETPAKRVVTVNHVDEGNPVAPOEGTKPKRSIPGEYFGTVDDEGNT 1136
 QY 814 GH 815
 DB 1137 TH 1138

 RESULT 7
 AB1_MOUSE STANDARD; PRT; 1321 AA.
 ID AB1_MOUSE
 AC 090Y30; 090Z8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Bile salt export pump (Atp-binding cassette, sub-family B, member 11)
 DE (Sister of P-glycoprotein).
 GN ABCB11 OR BSEP OR SPGP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=20076398; Pubmed=10607905;
 RA Green R.M., Hoda F., Ward K.L.;
 RT "Molecular cloning and characterization of the murine bile salt export pump.";
 RT Gene 241:117-123(2000).
 RN [2]
 RP SEQUENCE OF 463-635 FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RA Salkar R., Suchy F.J., Ananthanarayanan M.;
 RT "Molecular cloning of mouse liver bile salt export pump (bsep).";
 RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS
 CC INTO THE CANALICULUS OF HEPATOCYTES.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN
 CC THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR IN
 CC MICROVILLI AND TO SUBRANULAR VESICLES OF THE HEPATOCYTES BY IN
 CC SITU.
 CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN

[illegible]

QY 137 GKAOLNEELTRTREGPAEVH-ALMDNGHFANDPPRRMNGIYPMNNKEPRLPGNHATYGCV 195S
| : : : : :
Db 130 GLVDINSEVKEIFSGIYAGVGAVLLTGF-----QIRLM-----V 164S

QY 196 ATAAAOIMR----YHSWPLGEG-GEFDYHAGSLVGNMSGTFGEKMYDWINPGNPDDLNTL 250S
| | | | | : : : : :
Db 165 TTGAQOIKRMKRFFYRIIRIMEIGFDC--TSVGEUNSPESD-----DINKID 210S
| : : : : :
QY 251 QSOVDAYATILMRDVSASYS---MSFYENGSGTASYVV-----GALRNPFYKRSQJLH 30LS
| : : : : :
Db 211 EAIADQMMLFLQRSTALSGLLTGYRKWKTLTVILVASPLIGTGAANVIGLSVAKKFELE 270S

ID	E13B_BACCI	STANDARD:	PRT:	682 AA.
AC	P23903;			
DT	01-NOV-1991 (Rel. 20, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Glucan endo-1,3-beta-glucosidase A1 precursor (EC 3.2.1.39) ((1->3)-beta-glucan endohydrolase) ((1->3)-beta-glucanase A1).			
GN	GICA.			
OS	Bacillus circulans.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1397;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 39-52.			
RC	STRAIN-WL-12;			
RX	MEDLINE=90185240; PubMed=2311931.			
RA	Yahata N., Watanabe T., Nakamura Y., Yamamoto Y., Kamimura S., Tanaka H.;			
RT	"Structure of the gene encoding beta-1,3-glucanase A1 of Bacillus circulans WL-12.";			
RL	Gene 86:113-117(1990).			
CC	- FUNCTION: LYSIS OF CELLULAR WALLS CONTAINING BETA-1,3-GLUCANS. IMPLICATED IN THE DEFENSE AGAINST FUNGAL PATHOGENS.			
CC	- CATALYTIC ACTIVITY: Hydrolysis of 1,3-beta-D-glucosidic linkages in 1,3-beta-D-glucans.			
CC	- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way			

CC	modified and this statement is not removed. usage by and for commercial
CC	entitles requires a license agreement (see http://www.isb-slb.ch/announce/
CC	or send an email to license@isb-slb.ch).
CC	-----
DR	EMBL; M34503; AAA2474.1; -
DR	PIR; J00420; J00420.
DR	InterPro: IPR000757; Glyco_hydro_16.
DR	Pfam; PF00722; Glyco_hydro_16; 1.
DR	ProSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
DR	Cell wall; Hydrolase; Glycosylase; Signal.
FT	SIGNAL
FT	CHAIN 1 38
FT	ACT_SITE 39 682 GLUCAN ENDO-1,3-BETA-GLUCOSIDASE A1.
FT	ACT_SITE 552 552 NUCLEOPHILE (BY SIMILARITY).
FT	ACT_SITE 557 557 PROTON DONOR (BY SIMILARITY).
SO	SEQUENCE 682 AA: 75465 MW: 864F407E34D4AD5 CRC64;
Query Match	3.2%; Score 142; DB 1; Length 682;
Best Local Similarity	19.8%; Pred. No. 0.29;
Matches 114;	Conservative 82; Mismatches 197; Indels 184; Gaps 32;
DB	-----
308	SOEHHDMIRGELASGRPYVY-AGNNQSIGAFVCDGYASDGTFRHNG-WG-GVSNGEYK 364
78	SATMND-----YSDVGAVKKGNNMVIDQAGYIYNMGHMSDGFNGYV- 125
365	LTLSPTSLSIGGEGIGF-----TYQDI-----TGIEPATPRAEGDALPILAKNIE 415
126	FTLSATTEIQLYSKANGKLEYQLFQFQINKTTITPAMPPTGCP-----OIT 171
416	AEYKSEGLANGVSIYNTGEEQSNLDLGYRLNKADGEVIEYKTSIN-ISWYGEHPES 474
172	ASFLGAGFT--YTFNNDSAVT-----YEAADLKLKYVKRYVNSSMHDIDNNAS 221
475	FSLAPNOLSGCINITTLIRRTGTEQWEPRVRAAGGYVNSIKVNTDNNVVVTVDNNG 534
222	GMIVDHFNGQ-----FTDGGGGYFNVTES---INVKLESKTSANLYITTFNEP 269
535	KLSTVPSFVADLNSYEHSTTGVQENSQSPDIRPVPALSTGATADVISIGVMAVY 594
270	-----TRNSYV--IIPTEGTTTPADANG-----SIGLPKRI 299
595	PGSSSNTPVWVSKDVLTLSEGDYTLWFRFSINNOKDEMKIGSVKTPTEYTHPLEFG 654
300	DGGA--PI--AKEL-----GNFY--YQININGQ--W-----VDSLSSQSKFYANG 339
655	HNQIST-----YTLDMANH---RVLPDTLKNLGLPENG-----LVVFERQOSS 697
340	YNNNSDANQMGYMDVYIGLWFOPIQENMQIRIGPLNGQAGNIGNFNVTYFIGNENA 399
698	SGSLMAQOEYHIK-----QGEFFYKPKPVYEGEIPD-----GSY----- 731
400	PRPDVSDQEDSISGTPDPDAIAGMNLWMQDEFNGTTIDTSKWNVETGYLLNNDPATWG 459
732	RATLHAFNQGQOOLYIK-GKRNTYVKIYNGVAVEAIESSEIRVPPNPARDVETISAPCI 790
460	NAELQHYNSQNYVVOGKLN-----IKAMDSKSPQDPNRYAQS----- 502
791	POETSIILFDISGKIYKNSLSAGHRADV-SRLPNG 826
503	-----SGKINTKDKLSIKYGRVDFRAKLPPTG 528
RESULT 9	
KEX2_YEAST	
KEX2_YEAST	STANDARD; PRT; 814 AA.
AC	P13134;
DT	01-JAN-1990 (Rel. 13, Created)
DT	01-JAN-1990 (Rel. 13, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Kexin precursor (EC 3.4.21.61) (KEX2 protease) (Proteinase yscf).
GN	KEX2 OR OPS1 OR YNL238W OR N1122.
OS	Saccharomyces cerevisiae (Baker's yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
MEDLINE=89025851; PubMed=2845974;
Mizuno K., Nakamura T., Ohshima T., Tanaka S., Matsuo H.;
"Yeast Kex2 genes encodes an endopeptidase homologous to
subtilisin-like serine proteases";
Biochem. Biophys. Res. Commun. 156:246-254(1988).
[2]
SEQUENCE FROM N.A.
MEDLINE=89160792; PubMed=2646633;
Fuller R.S., Brake A., Thorner J.;
"Yeast prothormone processing enzyme (Kex2 gene product) is a Ca2+-
dependent serine protease";
Proc. Natl. Acad. Sci. U.S.A. 86:1434-1438(1989).
[3]
SEQUENCE FROM N.A.
MEDLINE=97051596; PubMed=8896273;
Pandofo D., de Antoni A., Lanfranchi G., Valle G.;
"The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open
reading frames including a novel gene encoding a globin-like
domain";
Yeast 12:1071-1076(1996).
[4]
PROCESSING, AND CHARACTERIZATION.
MEDLINE=92141230; PubMed=1736507;
Brenner C., Fuller R.S.;
"Structural and enzymatic characterization of a purified prothormone-
processing enzyme: secreted, soluble Kex2 protease";
Proc. Natl. Acad. Sci. U.S.A. 89:922-926(1992).
[5]
PROCESSING.
MEDLINE=92183954; PubMed=1544507;
Germain D., Dumas F., Vernet T., Bourbonnais Y., Thomas D.Y.,
Boileau G.;
"The pro-region of the Kex2 endoprotease of *Saccharomyces cerevisiae*
is removed by self-processing";
FEBS Lett. 259:283-286(1992).
[6]
POST-TRANSLATIONAL MODIFICATIONS.
MEDLINE=92011876; PubMed=1918142;
Wilcox C.A., Fuller R.S.;
"Posttranslational processing of the prothormone-cleaving Kex2
protease in the *Saccharomyces cerevisiae* secretory pathway";
J. Cell Biol. 115:297-307(1991).
[7]
FUNCTION: PROCESSING OF PRECURSORS OF ALPHA-FACTORS AND KILLER
TOXIN.
[8]
CATALYTIC ACTIVITY: Cleavage of Lys-Arg-|-Xaa and Arg-Arg-|-Xaa
bonds to process yeast alpha-factor pheromone and killer toxin
precursors.
[9]
COFACTOR: CALCIUM DEPENDENT.
[10]
SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. LATE GOLGI
COMPARTMENT.
[11]
PTM: O-GLYCOSYLATED.
[12]
SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8, PURIN SUBFAMILY.
[13]
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).
or send an email to license@isb-sib.ch).
[14]
EMBL; M22870; AAA34719.1; -
EMBL; M24201; AAA34718.1; -
EMBL; Z69381; CA93360.1; -
EMBL; Z71514; CA936143.1; -
PIR; A28931; KXBY.
PIR; S42157; S42157.
HSSP; P04072; 1THM.
MEROPS; S08.070; -;
SGD; S0005182; KEX2.

Page 9

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

DR EMBL; AF312024; AAG48303.1; -
 DR EMBL; AY010111; AAG27034.2; -
 DR EMBL; AB053445; BAB61902.1; -
 DR HSSP; P1516; INCJ.
 DR Gene; HGNC:13733; CDR23.
 DR MIM; 605516; -
 DR MIM; 601067; -
 DR InterPro; IPR002126; Cadherin.
 DR Pfam; PF00028; cadherin; 47.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 26.
 DR PROSITE; PS00232; CADHERIN_1; 17.
 DR PROSITE; PS50268; CADHERIN_2; 27.
 Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 Signal; Multigene family; Alternative splicing; Vision; Polymorphism;
 Disease mutation; Usher syndrome; Deafness; Retinitis pigmentosa.

FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 3354 CADHERIN 23.
 FT TRANSMEM 3065 3085 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 3086 3354 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 34 132 CADHERIN 1.
 FT DOMAIN 133 236 CADHERIN 2.
 FT DOMAIN 237 348 CADHERIN 3.
 FT DOMAIN 349 460 CADHERIN 4.
 FT DOMAIN 461 561 CADHERIN 5.
 FT DOMAIN 562 671 CADHERIN 6.
 FT DOMAIN 672 784 CADHERIN 7.
 FT DOMAIN 779 890 CADHERIN 8.
 FT DOMAIN 891 995 CADHERIN 9.
 FT DOMAIN 996 1102 CADHERIN 10.
 FT DOMAIN 1103 1208 CADHERIN 11.
 FT DOMAIN 1210 1313 CADHERIN 12.
 FT DOMAIN 1314 1418 CADHERIN 13.
 FT DOMAIN 1420 1527 CADHERIN 14.
 FT DOMAIN 1529 1634 CADHERIN 15.
 FT DOMAIN 1635 1744 CADHERIN 16.
 FT DOMAIN 1745 1851 CADHERIN 17.
 FT DOMAIN 1852 1959 CADHERIN 18.
 FT DOMAIN 1960 2069 CADHERIN 19.
 FT DOMAIN 2070 2174 CADHERIN 20.
 FT DOMAIN 2175 2293 CADHERIN 21.
 FT DOMAIN 2297 2402 CADHERIN 22.
 FT DOMAIN 2403 2509 CADHERIN 23.
 FT DOMAIN 2510 2611 CADHERIN 24.
 FT DOMAIN 2614 2722 CADHERIN 25.
 FT DOMAIN 2729 2846 CADHERIN 26.
 FT DOMAIN 2847 2975 CADHERIN 27.
 FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 694 694 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 765 765 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 810 810 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 941 941 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1001 1001 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1018 1018 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1171 1171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1282 1282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1315 1315 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1473 1473 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1651 1651 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1657 1657 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1818 1818 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1857 1857 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1889 1889 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1902 1902 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2013 2013 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2050 2050 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2129 2129 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2168 2168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2195 2195 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2263 2263 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2357 2357 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2369 2369 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2616 2616 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2749 2749 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2808 2808 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2877 2877 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2896 2896 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2941 2941 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2981 2981 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 379 380 MISSING (IN ISOFORM 2).
 FT VARSPLIC 1403 1403 MISSING (IN ISOFORM 3).
 FT VARSPLIC 3212 3212 MISSING (IN ISOFORM 4).
 FT VARSPLIC 3 3 R -> C.
 FT VARSPLIC 490 490 /FTid-VAR_012166.
 FT VARSPLIC 496 496 G -> A.
 FT VARSPLIC 496 496 S -> N.
 FT VARSPLIC 990 990 /FTid-VAR_012168.
 FT VARSPLIC 990 990 D -> N (IN DEFBI2).
 FT VARSPLIC 1222 1222 /FTid-VAR_012169.
 FT VARSPLIC 1222 1222 A -> T.
 FT VARSPLIC 1281 1281 /FTid-VAR_012170.
 FT VARSPLIC 1281 1281 MISSING (IN USH1D).
 FT VARSPLIC 1349 1349 /FTid-VAR_012171.
 FT VARSPLIC 1349 1349 R -> C.
 FT VARSPLIC 1351 1351 /FTid-VAR_012172.
 FT VARSPLIC 1351 1351 D -> N.
 FT VARSPLIC 1496 1496 /FTid-VAR_012173.
 FT VARSPLIC 1496 1496 O -> H (IN USH1D).
 FT VARSPLIC 1575 1575 /FTid-VAR_012174.
 FT VARSPLIC 1575 1575 T -> A.
 FT VARSPLIC 1671 1671 /FTid-VAR_012175.
 FT VARSPLIC 1671 1671 T -> S.
 FT VARSPLIC 1675 1675 /FTid-VAR_012176.
 FT VARSPLIC 1675 1675 V -> I.
 FT VARSPLIC 1999 1999 /FTid-VAR_012177.
 FT VARSPLIC 1999 1999 R -> O (IN USH1D; MILD RETINAL AFFECTON).
 FT VARSPLIC 2044 2044 /FTid-VAR_012178.
 FT VARSPLIC 2044 2044 /FTid-VAR_012179.
 FT VARSPLIC 2044 2044 R -> O.
 FT VARSPLIC 2045 2045 /FTid-VAR_012180.
 FT VARSPLIC 2045 2045 E -> K.
 FT VARSPLIC 2045 2045 /FTid-VAR_012181.
 FT VARSPLIC 2045 2045 D -> N (IN DEFBI2).

Query Match 3.1%; Score 136.5; DB 1; Length 3354;
 Best Local Similarity 20.1%; Pred. No. 6.3;
 Matches 199; Conservative 126; Mismatches 338; Indels 327; Gaps 52;

QY 21 SAPYTKERALSIALALROVSLRMO-TAVSDKISIDYVYRQGAERGTSQEE---GS 75
 DB 344 NAFENSSSESVATTELAQGFALPLFQVYDX-----DENGLNSMEFVLYGN 393
 QY 76 PAYTYVVRNNGEGYALVAADRIPTILASPIGRFMD-----SMEDNL---RMTLQTY 127
 DB 394 NSHFIIISPSVQKADIRIRVAIP--LDVETVRYDFDLFRANSVDHYGAYAKITLI 451
 QY 128 DQELGLISGAQNGEELIRTEGVPAEVHLM--DNG-----HFANDPMKNGGYP 176


```

Db 452 NENDNRPISQPLNYSLSENVTVGTSVLVLTADNDAGTGGVSYFESDDPREFSL--- 508
QY 177 MNKKPELPLPGNANAYGCVATAA-----AQIMRHSPLOCESGFDYHAGSLVGNMSTFG 232
Db 509 --DKD-----TGLIMLIARLDYELLORFTLTIIARDGGEEETGVRIN----- 550
QY 233 EMTDWIMPGNPDLDNLTOVDAYATLMRDVSASYSM-----SFEKNGSGTYSYV 283
Db 551 -VLD-VN-----DNVPTFOKDAYVGALEARENPSYTOVLRLATDEDSPPNOITYSIY 601
QY 284 VVGALEARENPSYTOVLRLATDEDSPPNOITYSIY 601
Db 602 SASAFSGSYFDI-----SLYEGY-----GVLSVRHLDYE----- 630
QY 344 ASAGTGHFMKMGNGCVNGFYLKLLSPSLGIGEGIFTIYOEITIGIPAKTPA----- 399
Db 631 -----QISNGLIYLYTV-----AMDAGNPLNSTVPVPTIEVFENDNPPTFSKP 674
QY 400 -----EAGTDALPLIALK-DIEAEKSSGSLVNGSYIYNTGEGSNLDLGRYLNK 448
Db 675 AYPVSVENIMAGATYLFILNATDLDSREYGOES-----IISL--EGSTO-----FRINA 723
QY 449 ADGEVY-----EVKTSINISWY---GYGEHPESESFLAPNQLSOGINT--ITLLYRRT 496
Db 724 RSGEIITTSILDRETSSEYILIVRAVDGCVGHNOKT-----GIATVNTITLDIND 773
QY 497 GTEGMEPVRAHAGGYNSIKVNTDPN-NVYVTVVDNNECK-----LSIYNSVADLNSY 550
Db 774 NHPTMDAPY-----YINLVEMTPDSDVTVVAVDLGENGLTVYSIOPPNFYSINST 829
QY 551 --EHSTITVOFNSDSD-----EIRTPVAFALSTGATADVISL----- 587
Db 830 TGRIRTHAMLDRENDDPHEALMRKIVSVTDCGRPLKATSSAVFVNLDLNDNDPT 889
QY 588 -----GMYMAVPGSSNYPYVNSKDVLLTSEGDTTLTMRSSINNOKDEMKGISVS 640
Db 890 FQNLPPVAEYLEBIPAGVSIYOV-----AIDDEG-----LN-----GLVSY 927
QY 641 KPTETHTPLFEVGHNOTSTYTLDMAHNRVLPDTLK-----NLGLPENGELVYVFPQOTS 966
Db 928 RMPVGRPRMDFLINSSSGVYVTTTELDREHIAEYOLRVASDAGTP-----TIS 976
QY 697 SSGSLMAADETVHI-----KOGETF---VYKPVVEGRIP-----DGSYRATLH 736
Db 977 SHTSL-----TIHVLVDNDETPFEFPAYVNVSEVDPREFRVWMLNCTDNDVGLMAELS 1031
QY 737 AYPVNG--QOOLYLKGRNTYKLVNG-----TA-----VEAISSEIRVFPNPARDYV- 783
Db 1032 YFTTGGVNDKESVGYRDAVVRVVGVLDBRETPAAYMLILEAIDNGFVGKRTGTATVFTV 1091
QY 784 -----EISAP-CIPEOTSTIIFDLDS-----GKI--VMKNSLSLASHG---R 817
Db 1092 VLDVNDNRPIFLQSSFEASVPEDIPGCHSILQKATADAGEGRVWYRILGNHGNR 1151
QY 818 MDVS-----RLPNGAYILKVDGY 835
Db 1152 IHVSNGLMRGPRPLDERRNSHVILVEAY 1181

```

RESULT 13

```

ID GUNZ_ERWCH STANDARD; PRT; 426 AA.
AC P07103;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase Z precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase Z)
DE (Cellulase Z) (Egz).
DE CELZ OR CEL5.
OS Erwinia chrysanthemi.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
OC Pectobacterium.

```

```

OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3937;
RX MEDLINE=88216177; PubMed=2835589;
RA Gillespie A., Caml B., Aymeric J.-L., Ball G., Creuzet N.;
RT "Homology between endoglucanase Z of Erwinia chrysanthemi and
endoglucanase of Bacillus subtilis and alkalophilic Bacillus.";
RL Mol. Microbiol. 2:159-164(1988).
RN [2]
RP REVISIONS, AND DISULFIDE BOND.
RC STRAIN=3937;
RX MEDLINE=94203057; PubMed=8152378;
RA Bortoli-German I., Brun E., Py B., Chippaux M., Barras F.;
RT "Periplasmic disulphide bond formation is essential for cellulase
secretion by the plant pathogen Erwinia chrysanthemi.";
RL Mol. Microbiol. 11:545-553(1994).
RN [3]
RP MUTAGENESIS, AND DOMAINS.
RX MEDLINE=91312880; PubMed=1677466;
RA Py B., Bortoli-German I., Hachez J., Chippaux M., Barras F.;
RT "Cellulase EgZ of Erwinia chrysanthemi: structural organization and
importance of His98 and Glu133 residues for catalysis.";
RL Protein Eng. 4:325-333(1991).
RN [4]
RP STEREOCHEMISTRY OF THE REACTION.
RX MEDLINE=92225124; PubMed=1563515;
RA Barras F., Bortoli-German I., Bauman M., Rouvier J., Gey C.,
Heyraud A., Henrissat B.;
RT "Stereochemistry of the hydrolysis reaction catalyzed by
endoglucanase Z from Erwinia chrysanthemi.";
RL FEBS Lett. 300:145-148(1992).
RN [5]
RP STRUCTURE BY NMR OF 365-426.
RX MEDLINE=98070232; PubMed=9405041;
RA Brun E., Moriana F., Gans P., Blackledge M.J., Barras F.,
Marion D.;
RT "Solution structure of the cellulose-binding domain of the
endoglucanase Z secreted by Erwinia chrysanthemi.";
RL Biochemistry 36:16074-16086(1997).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.30 ANGSTROMS) OF 44-335.
RX MEDLINE=21392910; PubMed=11501995;
RA Chapon V., Czjzek M., El Hassouni M., Py B., Juy M., Barras F.;
RT "Type II protein secretion in Gram-negative pathogenic bacteria: the
study of the structure/secretion relationships of the cellulase cels
(formerly EgZ) from Erwinia chrysanthemi.";
RL J. Mol. Biol. 310:1055-1066(2001).
CC -1- FUNCTION: REPRESENTS 97% OF THE GLOBAL CELLULASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DB EMBL: Y00540; CAA68604.1; -
DB PIR: S03767; S03767
DB PDB: 1ATW; 06-MAR-98.
DB PDB: 1EGZ; 26-MAR-99.
DB InterPro: IPR003610; CBM_5_12.
DB InterPro: IPR001547; GH_5.
DB Pfam: PF00150; cellulase; 1.
DB Pfam: PF02839; CBM_5_12; 1.
DB SMART: SM00495; ChIB3; 1.
DB PROSITE: PS00659; GLYCOSTYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; signal; 3D-structure.

```


ID YD94_METJA STANDARD: PRT; 987 AA.
 AC Q58789;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein MJ1394.
 GN MJ1394.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Sutton G.G., Blake J.A., Olsen G.J., Zhou L., Fleischmann R.D.,
 Kertavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Gocayne J.D.,
 Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
 Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 Cotton M.D., Roberts K.M., Hurst M.A., Kalne B.P., Borodovsky M.,
 Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii"; Science 273:1058-1073(1996).
 RL Science 273:1058-1073(1996).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: TO M.JANNASCHII MJ1393 AND A.FULGIDUS AF2028
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U67579; AAB99404.1; -
 DR TIGR: MJ1394; -
 KM Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 12 32 POTENTIAL.
 FT TRANSMEM 958 978 POTENTIAL.
 SQ SEQUENCE 987 AA; 112360 MW; D1E628FB28CA86D CRC64;
 Query Match 3.08; Score 135; DB 1; Length 987;
 Best Local Similarity 18.68; Pred. NO.1.3;
 Matches 155; Conservative 94; Mismatches 293; Indels 292; Gaps 42;
 72 EEGSPAFVYANRGN-----NEGAL-----VAADDRIPRLASPIGRPMDSMPDN 119
 295 ESSAPAVTNLPNANTYIHILPNNSTVEYEDIDASQTLPIIVNET---DYVTKIPAN 350
 120 LRMLDIYDOEIGLISGRAQLNEEL-RTEGVPAEVALMDGHPANDPMRNGYPMN 178
 351 -----KKEWTNLIY-----YLKNLVPNGENVNVVITYLSGQPYDNEFKMT---GNM 398
 179 N-KEPL-----PNCNHAY-----TGCVAATAAQIMRYS 207
 399 QIKNGIVQWSSIQSHSGNYSLEKYGISTSLNDPNNGYKILPKELGRDVIYSGWVYRPSN 458
 208 WPLQGGSEFDYHAGSLVGNMSTFGEMVDMIN-----MPCNPLDNLTOGVNAYATL 260
 459 W---GGGPID-RIGLEDENFDGSEFVNYSYISTDRTPNTP-----TEISPEYVNP 509
 261 MNDVASVMSSEYENGSGTYSVYVGALNNRFRYR-SLQLHVALYLTQSEWHDIMRGL 319
 510 PEDWEYFELKIKYISGTITFTSY-----YNGSLAATVSTIDNTYKFDV---V 556
 320 ASGRVYVYAGNNSIGHAFCYASDGTFFH-----NMGWGVSNGFYKLLSLPTSLGI 375
 557 IHGCVYVY-----VVDLEVNSKNFDFYGDKNKY----- 585
 376 GCEGICGTIYOEITINGIEPAKTPAPAGTALPILAKDIEAEYKSESGLANVGYSTYNGE 435

DB 586 -----LETSANSSECTAVL-----FDGDFKKD-----YNT-- 612
 QY 436 EOSNLIDGYRLNKADGEVIEVTKSSINISW-----YGEHPESEFSLAPNG 481
 DB 613 ---SNLN-----AINNTNITLMSNSATLVEVLCNYSERDNILA----- 652
 QY 482 LSGCINTYLLYRRGTGEOMEPEVHAOGGVNSIKVNTDPNNV---VYVDNNEKLSI 538
 DB 653 -KYGFALIFENYNGTNTNITSIKGVYASGSYSTDGHGTGTEINIEWNTFKNDASYSF 711
 QY 539 -VPSFADLNSYENSHITVQENS-----DSPDEIRTPVAFALSTGATADVISLGVMAE 593
 DB 712 NLNLNIMAVNKAKEYELYNPNPKSJWIDGNTTTP-----NID 751
 QY 594 VPGGS-----SNPVVWSKDVLTISEGDYTLMYRFSINNOKDEMKRIGSVYKT 642
 DB 752 IPGGEVWNSKTYNFTSGVPIVWANGSPILSKDYL-----LNEVSQIGSSYV-- 800
 QY 643 PTETTPLEPEVGNHNTSTYTLDMANRVLDP-----FTLKNIGLPENGELVYVF-- 691
 DB 801 ---VVEIIVYG-----STLIKVT-KHIVPDADGTVDIYIVENIGSVTPPEVYVYDIL 851
 QY 692 --ROGSSGSLMAAQETVHIKOG-ETFYKRP-----VVEGPIPGGSYRATLH 736
 DB 852 PKNFTVSD-----WVNOSSMLIAEGNHTTTNPRTNYSMMWALHAIYPGADGGMNDYAE 908
 QY 737 AFVNGOOLYLKGRNRYTVKIVNGTAVEAIESEETRPVNPARDVETAPCI 790
 DB 909 ILANKTVVIHYK-----LNGTG--EFYPSDAFVIGIDPTNSLPTTSKRI 951
 RESULT 16
 AMY_BUTFI STANDARD: PRT; 976 AA.
 ID AMY_BUTFI
 AC P30269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alpha-amylase precursor (EC 3.2.1.1) (1,4-alpha-D-glucan
 DE glucanohydrolase).
 GN AMYA.
 OS Butyrlivibrio fibrisolvens.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
 OC Butyrlivibrio.
 OX NCBI_TaxID=831;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H17C;
 RX MEDLINE=91286207; PubMed=2061294;
 RA Rumbak E., Rawlings D.E., Lindsey G.G., Woods D.R.;
 RT "Cloning, nucleotide sequence, and enzymatic characterization of an
 RT alpha-amylase from the ruminal bacterium Butyrlivibrio fibrisolvens
 RT H17C";
 RL J. Bacteriol. 173:4203-4211(1991).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 CC linkages in oligosaccharides and polysaccharides
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M62507; AAA23005.1; -
 DR PIR: A42466; A42466.
 DR HSSP: P00691; 1BAG.
 DR InterPro: IPR000461; Alpha_amylase.
 DR InterPro: IPR005085; CBM_25.

DR Pfam: PF00128; alpha-amylase; 1.
 DR Pfam: PF02806; alpha-amylase_C; 1.
 DR Pfam: PF03423; CBM25; 1.
 KM Hydroxylase; Glycosidase; Carbohydrate metabolism; Signal.
 FT SIGNAL 1 33
 FT CHAIN 34 976
 FT ACT_SITE 323 323
 FT ACT_SITE 327 327
 FT ACT_SITE 447 447
 SQ SEQUENCE 976 AA; 106695 MW; 610690EDC19315AC CRC64;

Query Match 3.0%; Score 134.5; DB 1; Length 976;
 Best Local Similarity 19.4%; Pred. No. 1.4;
 Matches 155; Conservative 116; Mismatches 259; Indels 267; Gaps 47;

QY 13 FGIN-----HOGSAPYTKERALSRL-----ALROVSLRMGTAVSDKISIDYVYRQG 62
 DB 232 YGVAIVIDILPNHTPTGTSTAKALMEAGGSDALYHATKIG-GGYTDRLLEY-YSMG 289
 63 DAERGTSOEGSPAYFYVANRG-----NNEGALVAA-----DDRPTILAYSPIGRFD 112
 290 GLP-DVTEMTGFOQYTFEFLKDCVYLGADGFRIDTKHISLPDDPPVS--DYSAGR-- 344
 QY 113 MDSMPDLRMWLOIYDDEIGLISGAQLNEELRTGVPALVADNGHFANDPMKRN 172
 DB 345 -NFEYPMNRALNEYSEVEGTSYDELFEVYGEVLOGTNDRLAAVQYIGGTAS-----N 398
 QY 173 QGYPMNKEPLPFGNIAVYGCATAAQIMRHSWPLQSGSFYDIAHSLVGMWSGTFG 232
 DB 399 YG-----SLRSALSGN-----LSVRLDYOIYDDTAAGS-----TYTADTE 437
 QY 233 EYDWMINPGN--PDLNLQSOVDAYATLM-----RDVASVMSFYENGSG--GYTS 281
 DB 438 KLYTWVSHDNYMDSCKKS-IDDDVIMGSIITAKRGCTLFSTRPNSSAENPYG 496
 QY 282 VYVVGALR-----NFRYK-----RSIQLVHRYALYTSQEHDM 314
 DB 497 DNLIGAAGSPIYKAPEYKAVLFRKKGEADEYLSPNGNIGTQIMT-----ERYNDT 548
 QY 315 IRC-----ELASG-RPYVYAGNNOISGAFVCDGAS----- 345
 DB 549 VQGVAVIYMAAOTRTTISTETHLSGDIYPDQVSGSNVYF--LVKDGVLSGSVEGEGVYL 605
 QY 346 ----DGT-----FHFNMGGV-----SNGFYKLTLLSPTSLGIGRGI 380
 DB 606 SEKMDGTGKYVSFTNNKNGVAVRVDAEETLTIDENCGEFOYTYLDE----- 656
 QY 381 GFTIYOEITIGIEPA---KTPAEGTDALPILALKDIEAEY--KSSGLNV-GYSIY-- 431
 DB 657 -FTIRFESADCKEVSPEFQITAEAGTATP---DSSELYSKAEAEBGIGIHTYVYFF 711
 QY 432 ---NTGEOGN--LDLGRILKADGEVIEKTSINISWIGYGHSPESFSLAPMOLSOGI 486
 DB 712 NTEMMGVSYYTGGDGAQLFGMPGTVAVANEGS--GWY-----RADKTTGEI 758
 QY 487 NTITLLRRTGTEQMEVVRHAGGVNSIKVN--TTDPNNVYTVVONNEGSLIVPSFV 544
 DB 759 TAPULIIN-----NGNGIOTVNEGITPTPSKDIYLVDAEKSGOQLIVNY- 804
 QY 545 ADLINSYEHSS-----TITVOFNSDPELITPAPALSTGATA-----DDVI 585
 DB 805 EDRKSAKALGVSGSYTATFAYNTEGMDK-----CAVTCATATLGDMPGKELTQDD- 857
 QY 586 SLGVMAAEVPEG--SSNPVNVSKDVLTLSEGYT-----LWTFRSINN-----Q 628
 DB 858 --GVYSVVLPAQPEEDNIITFNNG-----NNCKQTNDMKISDMKYRFTLLNNGISYQKXGSK 911
 QY 629 KDEWKI---GSYSVKT 642
 DB 912 KDAMEALAGAGDYET 928

RESULT 17

HEC1_HAEIN
 ID HEC1_HAEIN STANDARD; PRT; 837 AA.
 AC P33397;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Outer membrane usher protein hifc precursor.
 GN HIFC.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 OX NCBI_TaxID:727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bagan / Serotype B;
 RX MEDLINE=94131580; PubMed=7905461;
 RA Watson W.J., Glisford J.R., Tucci M.A., McCrean K.W., Forney L.J.,
 RA Mairs C.F.;
 RT "Identification of a gene essential for pilation in Haemophilus
 RT influenzae type b with homology to the plus assembly platform genes
 RT of Gram-negative bacteria".
 RL Infect. Immun. 62:468-473(1994).
 CC -1- FUNCTION: ESSENTIAL FOR PILATION.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
 CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-97 IS THE INITIATOR.
 CC
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC
 DR EMBL: U02932; AAB53096.1;
 DR InterPro: IPR000015; FimB_usher.
 DR Pfam: PF00577; Usher; 1.
 DR PROSITE: PS01151; FIMBRIAL_USHER; 1.
 KW Outer membrane; Transmembrane; Fimbril; Transport; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 837
 FT DISULFID 813 833
 SQ SEQUENCE 837 AA; 92733 MW; BA529323AF63BF8B CRC64;

Query Match 3.0%; Score 133; DB 1; Length 837;
 Best Local Similarity 21.3%; Pred. No. 1.4;
 Matches 129; Conservative 80; Mismatches 212; Indels 186; Gaps 32;

QY 22 APYTKERALSRLALR-----QVSLRMGTAVSDKISIDYVYRQGDAREGITSQEG 74
 DB 282 APVVRGVAANTAKSISIKONGTYIQTVPAPFVINDLYASG---SGDLTVEI-QESDG 337
 QY 75 SPAVEYVNRGNNGGVALLVADDRIPITLAVSPGPFMDMPNLRLMWLOIYOEI--G 132
 DB 338 KYRSEIYV-----FSNLAPLMRYGHLRIVLAGGRIRIDS-----RTFERVLOG 381
 QY 133 LI---LSGKAQLNEELRTGVPALVADNGHFANDPM--RMNGYPMNKE--PLPNG 187
 DB 382 VLOGYGLNHLNLSLILTRYRAGLF-----GGLMTPIGAFAADATWSAEPLPKYVS 436
 QY 188 NHAYT-----GCVATAAQIMRHSWPL----- 210
 DB 437 KNGYSLGSYSINENEIGTNLTLLAA--YRYSRDFYTLSDTIGLRTFRQSGAYLPEIY 494
 QY 211 QGSGSFYHAGSLVGNMNGSF--GEYDWMIMPENPDLNLTOSOVDAVATLMDVASV 268
 DB 495 RPKNQFOVSLQSGISGNNGNTLYLSGQTYNWEKRTN-----TOYOV-AYSSEHILINYSV 548
 QY 269 SMS---EYENGSGTYVYVVGAL-----RNPFYKRS---QLVHRYALYTSQEHDM 314
 DB 549 NLSQSIDKFKCKRNSTIYLSLPLGDNHSDSSYSRSGNDINRLGVNSGF--GERHOW 606

OY 315 IRGELASGRPVYVYAGNNSIGHAFVCDYAS-DGTFHFMWGGVSGNGFKLTLS-PTS 372
DB 607 SYGINS-----RNNQ-----GRTSYDGNLSHNNSTISYASYSRDLKRNSTIS 650
OY 373 LGTGGGEGT---GFTYQELITIGIEPAKTPAEAGTALPLALKDIEAEYKSSGLNVGY 428
DB 651 LGVSGAVVAHKYGTITLSQPV-----GSEFALTHAKD--AAGAKVESGANVSL 695
OY 429 STYNTGEOSNLDLGRKNKADGEVIEKTSINISWGY-----GEHPESFSLAPNQLSQ 484
DB 696 DVF-----GNAAVPTTSPYEINYGINPSDAEANVEFEATEROILP 736
OY 485 GINTITLLYRTGTGEQMEPRHAQGGYVNSIKVNTDPNNVVY-----TYDNNKGKLTIVP 540
DB 737 RANISILVDFRICKN-----TMVLEFNLLPNEEPVPASTADDSGC----- 777
541 NSFVADL 547
778 -AFVGDV 783

RESULT 18
EAE_ECO57 STANDARD: PRT: 934 AA.
ID EAE_ECO57
AC P43261; 085627;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Intimin (Attaching and effecting protein) (Eae protein) (Gamma-
DE Intimin)
GN EAE OR EAEA OR 25110 OR ECS4559 OR L0025.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=92204008; PubMed=1552854;
RA Yu J., Kaper J.B.;
RT "Cloning and characterization of the eae gene of enterohaemorrhagic
RT Escherichia coli O157:H7";
RL Mol. Microbiol. 6:411-417(1992).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / CL-8;
RX MEDLINE=92249750; PubMed=1577255;
RA Beebakhee G., Louie M., de Azavedo J., Brunton J.;
RT Cloning and nucleotide sequence of the eae gene homologue from
RT enterohaemorrhagic Escherichia coli serotype O157:H7";
RL FEMS Microbiol. Lett. 70:63-68(1992).
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=98339885; PubMed=9673266;
RA Perna N.T., Mayhew G.F., Postal G., Elliott S., Donnenberg M.S.,
RA Kaper J.B., Blatter F.R.;
RT "Molecular evolution of a pathogenicity island from enterohaemorrhagic
RT Escherichia coli O157:H7";
RL Infect. Immun. 66:3810-3817(1998).
RN 14
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / DEC 3a, and O157:H- / DEC 3f;
RX MEDLINE=99261633; PubMed=10331248;
RA McGraw E.A., Li J., Selander R.K., Whittam T.S.;
RT "Molecular evolution and mosaic structure of alpha, beta, and gamma
RT Intimin of pathogenic Escherichia coli";
RL Mol. Biol. Evol. 16:12-22(1999).
RN 15
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dinalanta E.T., Potamotis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatter F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN 16
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo H., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kohara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
RN 17
RP SEQUENCE OF 1-37 FROM N.A.
RC STRAIN-O157:H7 / HAL;
RX MEDLINE=96079490; PubMed=8566710;
RA Zhao S., Mitchell S.E., Meng J., Doyle M.P., Kresovich S.;
RT "Cloning and nucleotide sequence of a gene upstream of the eae gene
RT of enterohaemorrhagic Escherichia coli O157:H7";
RL FEMS Microbiol. Lett. 133:35-39(1995).
CC -1- FUNCTION: NECESSARY FOR THE PRODUCTION OF ATTACHING AND EFFACING
CC LESIONS ON TISSUE CULTURE CELLS.
CC -1- SUBCELLULAR LOCATION: OUTER SURFACE.
CC -1- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 LYSM REPEAT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: Z11541; CAA77642.1; -;
DR EMBL: X60439; CAA42967.1; -;
DR EMBL: AF071034; AAC31504.1; -;
DR EMBL: AF081182; AAD05499.1; -;
DR EMBL: AF081183; AAD05499.1; -;
DR EMBL: AE005595; AAC58823.1; -;
DR EMBL: AE002566; BAB37982.1; -;
DR EMBL: U32312; AAB00111.1; -;
DR InterPro: IPR003344; B1g_1.
DR InterPro: IPR003343; B1g_2.
DR InterPro: IPR003535; Intimin.
DR InterPro: IPR002482; LySM.
DR Pfam: PF01476; LySM; 1.
DR Pfam: PF02368; B1g_2; 1.
DR Pfam: PF02369; B1g_1; 2.
DR PRINTS: PR01369; INTIMIN.
DR SMART: SM00257; LYSM; 1.
KW Outer membrane; Virulence; Complete proteome.
FT REPEAT 65 113
FT CONFLICT 221 221
FT CONFLICT 311 312
FT CONFLICT 318 318
FT CONFLICT 642 642
FT CONFLICT 769 769
FT CONFLICT 771 772
SQ SEQUENCE 934 AA; 101835 MW; BE73686D/F79D3C9 CRC64;
Query Match 3.0%; Score 133; DB 1; Length 934;
Best Local Similarity 17.2%; Pred. No. 1.6;
Matches 165; Conservative 128; Mismatches 338; Indels 326; Gaps 44;
OY 74 GSPAYFYVANR--GNNEGVALVAADRIPTILAYSPIGRF-----DMSMPD-N 119

```

Db      26 GIGLFFVYVQNSPANEENFKLGSKSLTHDSYQ--NRLFYTLKTGETVADLSKSDIN 83
Qy      120 L-RM---LQIYQELGLISGAQOLNEELRTGEPVAFVHALMDNCHPANDMRNNGY 175
Db      84 LSTIWSLNRHLVSESEMKAAAPQ--QITLPLKRLPEFSAL----- 124
Qy      176 PMNKEPRLPNC-----NHAUYGVATAAQAQIMRYHSMPLQG 212
Db      125 PILGSAPIYAGVAGVAGHTNKLTKMSPDVTKSNMTDDKALYAAQQAASLGSQLSRLNG 164
Qy      213 EGSFDYHAGSLVGNMSTGEGEMTDWIMNPGNPDILNTOQVDAVATLMRDVASVSMF 272
Db      185 DYAKDPAAG- IAGNOAAS- QLOAWLOHYGTAEVNLQSGNNPFG--SSIDFLPPE 234
Qy      273 YEN-----GSGT-----YSYV-----VGL 288
Db      235 YDEKMLAAGVQVARIYDSRFTANLGAQGFLLPANMLGYNVETDDESGDNTRLIGGE 294
Qy      289 RNNRYKRSLOLHVRLALYTSQEWMDMIRGELASGRPY-----YAGNNOISGHAFC 340
Db      295 YMDYFRSS---VNGYFRMSGWHESYNNKDYDERPANGDIRFNGILPSYPALGAKLIY 350
Qy      341 DGASDGTFRHNGMGVSNNGFYKLLLSFTSLGIGGEGIGFTIYOEITIGIEPAKTPAE 400
Db      351 EQYYGNVALFNSD-----KL-----QSNPGAATVGVNTPIPILVTMGI----- 389
Qy      401 AGTDALPILAKLIEAEKSESLANGYSTYNGEERS---NIDLGRILKAKAGEVIEVK 457
Db      390 -----DYS-----HGTGENEDLLYSMFRQFDRKSMQOIEPQ 422
Qy      458 TSSINISWYGYGHPESFLAPNQLSOGINTLTLRYR-----TGTEQ----- 500
Db      423 -----YVNEKLTSSRYDLYQRNNNITILEKKOILSLNPHDNGEHSQKI 472
Qy      501 -----W-EPVRAAGGYVNSIKVNTDPNNVYVVDNNEKGLSIYVNSEFA 545
Db      473 OLIVSKRYGLIRIYMDSDALSRSQGOIQHSQSASADYQALIPAYVQGS-----NIXV 527
Qy      546 DLSYSHS-----TIVQFNSDSPEIRTPVAFALSTGATADVISLGVMAVPE 596
Db      528 TARAYRNGNSSNNVOLTITVLSNGVVDVGVYDFTADTKSAKADADITTYATVYKKN 587
Qy      597 G--SSNYPVYMSKDVLTLSGDTLWYRFISNNOKDEMKIGSVYKPTREYHPLPEVG 654
Db      588 GVAQNAVPSFN-----IYSGTATL--GANSKAKDANGKATVTLKSSIP-----G 630
Qy      655 HNOTSTYTLDMAHNRVLPDFTLKNLGLFENGELVVFVROTQSSSGSLMAAQETVHIKQGE 714
Db      631 QVVVASAKTAEMTS-----ALNASAVIFEDQTKASITTEI-KADKTTAVANGK 675
Qy      715 ---TEYKKPVYEG-PIPDGSYR-ATLHAFVNGOQLYL-----KGRNY 753
Db      676 DAIKYVYVWKNGQPNOSVTPSTNFGMFGKSGQVQATTGNDGRATITLSSSAGKATV 735
Qy      754 TVYIVNGTAVEAIES--EIRV-----FPPARDYVEISAPC----- 789
Db      736 SATVSGAEKATKAEVFFDELKIDNKVDITGNVREGELPRIMVQYGFKLKASGGDCTYS 795
Qy      790 -IPQETSIILFDLSGKIYVKNLSLASHGRMDVSRPLNG-----AYILKVGYYTKIN 840
Db      796 WYSENTSIATVDSAGVYTLN-----GKGSY-VIKATSGDKQTVSYTIKASYSIMKYD 846

```

```

RESULT 19
CSD2_ECOLI
ID CSD2_ECOLI STANDARD; PRT; 802 AA.
AC P53513;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane usher protein csd precursor (Csd6 fimbria usher protein).

```

```

GN CSD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=88775;
RA Wolf M.K., de Haan L.A.M., Cassels F.C., Willshaw G.A.,
RA Gaestel E.C.M., Gaestel W., Warren R., Boedeker E.C.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF C6 FIMBRIAL
CC SUBUNITS ACROSS THE OUTER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; U04846; AAB51364.1; -.
DR InterPro; IPR000015; Fimb_usher.
DR Pfam; PF00577; Usher; 1.
DR PROSITE; PS01151; FIMBRIAL_USHER; FALSE_NEG.
DR KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT SIGNAL 1
FT CHAIN ? 802 OUTER MEMBRANE USHER PROTEIN CSD.
SQ SEQUENCE 802 AA; 90393 MW; 5BFDOB617F5B580 CRC64;

```

```

Query Match 3.08; Score 132.5; DB 1; Length 802;
Best local similarity 18.88; Pred. No. 1.4;
Matches 147; Conservative 102; Mismatches 275; Indels 259; Gaps 42;

```

```

Qy      133 LIISGKAQLNEEL-LTREGVPAVHLMONGHPRANDPMRNGQYPMNNEPILPNCNHA 190
Db      132 LFTAVPKNLSEIDKISKDIDNGIALFTN-----YQVNTN--LANKNR 175
Qy      191 YFGCVATAAQAQIMRYHSMPLQEGSEFDYHAGSLVGNMSTGTF-----GEM 234
Db      176 YD---YISVPNINYPFMRRL--NLFEFNONDEKTEMRNTYLEKSFYDKLNLVYGES 230
Qy      235 YDMINPQN-----PDLNITQSOVDAYATLMRDVASVSMSTYENG----- 276
Db      231 YTNSNYYNNYSFTGISVSTPDWYTPSEID-YTPEIIGVADSDQIIVROGNTIINESV 289
Qy      277 -SGTYSYVVYVVALRNNRYK-RSLQLHVRLALYTSQEWMDMIRGEL--ASGRPY----- 326
Db      290 PAGEFSPT-----TNLMYGGQLNVEITDIYGNKKQIYVNNSSLPMKRAGLVNYFIS 344
Qy      327 --YAGNNOISGHAFCVAGSDGTFRHN---WCGGVSNNGFYKLLLSFTSLGIGGEGIG 381
Db      345 GKLLKKNSDGD-FTFGDGLNIGYR-HYNSLFLGQYQSKNYFNLSITGIDLGSGAML- 401
Qy      382 FTIYOEITIGIEPAKTPAEAGTALPILAKLIEAEKSSGSLNVGS-----IYVGE 435
Db      402 -----LHVSRSNKKRNGYININIQONTQLPFPVAG- 431
Qy      436 EQSNDLGYRLNKADGVEIVKTSINISWYG--YGEHPSEFLAPNQLSOGIN---TIT 490
Db      432 --VNFDTAYKKR---YVELS---DIGWGNLYLNKLKNSFSL--LSKSLKYNFIS 478
Qy      491 ILYRRTGEQMEVPRHAAGGYVNSIKVNTDPNNVYVVDNNEKGLSIYVNSEFAVDNSY 550
Db      479 LDVNRK--MKYWD-----NAQDSNMSMIRYFFKPRAMITTNCLNKQSY 521
Qy      551 EHSITIVQFNSDP---DEIRTPVAFALSTGATADVIS-----LCW----- 589
Db      522 EKKDKRPSINISLPLTKDYGHISNYSFSNANGTATSSVGLNCSFENDARLMMNIQNR 581

```

QY 590 -----VMAEVPDSSNYP-----VWMSKD--VLTLSRG 615
 DB 582 TTRNNGYTNDNTSYATSYASPYGVFTGYSGNKYSOSSQFYASGGLVHSDGAFYOKAG 641
 QY 616 DYTLMATRESINNOCKEMKIGSVYKTPREYR--HPLFEVGHANQSTYTLDMANRYLPD 673
 DB 642 DTSALVR--IDNLS--IKIGN--TPGVYTGNGFALIPHLQPFKKWTILLDKGIDP 693
 QY 674 -FLKLKLG--LPENGELY-----VYFROTQSSGSLMAAQETVHIKGFETFY 718
 DB 694 GTTLANIKKOVIPSRKAIVKVPDAKKNDILFKLT-TYDGKTPPLGAIHAHERNG-- 747
 QY 719 KPYVEGPIPDGYSRATLHAFAVNGQQLYLKGRKNTYKIVNGTAVEAIESSEIRVFPMP 778
 DB 748 KQNTGIVDDG-----MLYMSG-----LSGTGIINTVMNGKVCSPFS 786
 779 ARD 781
 787 ERD 789

RESULT 20

VG37_BPARI STANDARD; PRT; 1103 AA.

AC 09G0B5;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Long tail fiber protein p37 (Protein Gp37) (Receptor recognizing protein).
 GN 37.
 OS Bacteriophage ARI.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 OC 74-like viruses.
 NC NCBL_TaxID=66711;
 RN 111
 RP SEQUENCE FROM N.A.
 RA MEDLINE-2048545; PubMed-11029414;
 RA Yu S.L., Ko K.L., Chen C.S., Chang Y.C., Syu W.J.;
 RT Characterization of the distal tail fiber locus and determination of the receptor for phage ARI, which specifically infects Escherichia coli O157:H7.
 RT J. Bacteriol. 182:5962-5968(2000).
 RL J. Bacteriol. 182:5962-5968(2000).
 CC -1- FUNCTION: STRUCTURAL COMPONENT OF THE DISTAL-HALF TAIL FIBER. IT CONSTITUTES THE PART OF THE LONG TAIL FIBERS THAT RECOGNIZES THE BACTERIAL RECEPTOR (BY SIMILARITY).
 CC -1- SUBUNIT: THE DISTAL HALF-FIBER CONTAINS TWO MOLECULES EACH OF GP36 AND GP37 AND ONE MOLECULE OF GP35 (BY SIMILARITY).
 CC -1- MISCELLANEOUS: THE TWO GP37 PROTEIN CHAINS RUN IN PARALLEL, THE LENGTH OF THE DISTAL HALF-FIBER, WITH THE AMINO END NEAR THE CENTER KINK OF THE FIBER AND THE CARBOXYL END AT THE DISTAL TIP. THE OTHER POLYPEPTIDES ARE DISTRIBUTED UNIFORMLY ALONG THE LENGTH OF THE DISTAL HALF-FIBER.
 CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: AF208841; AAG29754.1; -
 KW Fiber protein.
 SQ SEQUENCE 1103 AA; 119177 MW; D7496C81C1DB056E CRC64;

Query Match 3.0%; Score 132.5; DB 1; Length 1103;
 Best Local Similarity 19.8%; Pred. No. 2.2; Indels 235; Gaps 36;
 Matches 142; Conservative 84; Mismatches 255; Indels 235; Gaps 36;

QY 171 WNOGYPMNKEPLPNGNHAYTGCYAT-----AAQIMRYHSGWPLQEGSGFDYHAGSLVG 225

DB 68 YNOTGDYTLNGEFTQTGNFNLGIAVRYDIIAAGIMT-----EG-----GELIT 113
 QY 226 MNSGT-----EGEYND-----WIMPGNPDLDNTQSOYDAVATLMRPVSAVSKSEFENS 277
 DB 114 KSGGTAVHVFPHSDADEKGIITSPAN--DGLTQYVN--IRVRYKASSSESTFAFNGN 167
 QY 278 GTSYVYVVALRNNEFYKRSLOLHVRYALYTSOEMHDMINGELASGRPVYAGNN----- 331
 DB 168 GLESSPEV-----FGWK--SVSPVITNKV--ITKKYKDDVDIYSMDNNVPLSRI 215
 QY 332 -OSIGHAYVDGASDGTTH-----PMMGWG-----GYS 359
 DB 216 TTAINHLYRWNAVSGSIFHEVYKNDGITYWGGDADAYLWSTWGGIKSSHSISIGLT 275
 QY 360 NGFYKLTLSPTSLGSGGEGITTYOE-----IING-----IEPAKP-----AE 400
 DB 276 PGPKDYSLIGPSSIALGNDGTGFKWQDDYIYSVNNGTITFLPSSETISLKEFYAGYST 335
 QY 401 AGTD-----ALPILAKDIEAEYKSEGLNV-----GYSIYNTGEQSNLDGLYRL 446
 DB 336 NGTDLTPTPTENALATVYTHYDNNNAFGDGGTLGYOGGNGHHYFRKGTGINHNGSL 395
 QY 447 NKADGEVIEYKTSINI-----
 DB 396 LVTPG-IIDVIGSVNIDGRNNASTAFKNTTSSSSVDNMTISVWNTFTNPSEGRKN 454
 QY 464 -----SMYGEHPESEFSLAPNOLSGINTITLLYRRTGEOMEPEYRHAOG-----GY 511
 DB 455 VNEISDATSMATIQ-----LTTGEVKNVNG--SESSGVTAGNRKVHTTGELSSGA 506
 QY 512 VNSIKVNTDPNNVYTVDNNEKLSIVNSEFADLSYEHSTITVOFNSDSPDEIRPV 571
 DB 507 VNALRIWMDYG--VIFRSEGLHIITPLAEGEKNG-----DIGPLR--PE 549
 QY 572 APLALSG-----ATADVISLGVN--MAEVPDSSNYPVYWSKDYVLTLSGD----- 616
 DB 550 STALDGVKVIIPDLESSYNTFAANGYIKFAGHAGAGGYDIOYSCAAPIFOEIDDAVSK 609
 QY 617 YTLATRESINNOCKEMKIGSVYKTPREYR--HPLFEVGHANQSTYTLDMANRYLPDPT 675
 DB 610 YPIYKOKLNGKAAV-SIGT-EINSGTIVLHLKEDSGOGHTSRNADGTYN--PPDNY 665
 QY 676 LKNLG--LPENGELYVFERQOTQSSGSLMAAQETVHI-----KQGETFYVK 719
 DB 666 QVGGGEATIARNGNIFSDIMKFTTSAG-----ETINIDALATRVSKGDMTNGK 715

Search completed: May 29, 2003, 13:26:02
 Job time : 22 secs

00: E0 VB0CNDABPCTMCEBECNDAYEVYVAVDCUNVECVAYVAAVDDNDTMTTYAVVCRTCHREDVWDCWVND 118


```

Db 74 YKAEEREE-----ALFEVFNKRGKDGFLVLAADRFPEVIGYAFKGFHDAARD 123
Oy 119 NLRMMLQIYDOEIGLILSGKAQLENEELR---TEGVAEYHALMDNGHFANDPMNNGY 175
Db 124 NLRGWILKGYRELELAWMDGAEIPIDIRREKPTDRDPSSTAPLLEGEHNSDPLMDQ 183
Oy 176 PMNNKEPRLPNHAAVTGCVATAAQMIRYHWPLOGESSFEDYHAGSLVG---NMSGTEG 232
Db 184 PFHTLPLRLPSGOAATGCVATAMGOIMRHYKMPKASGEYDY--DDMTGTHHYSGTGFG 242
Oy 233 EMDWMINMGPNPLDMLTOSOVAYATLMMDVSAVSMSPYENGSGTSYUYVGAALRN 292
Db 243 ETNMSKMPGNISV--GISPEEVALSTFMDVSFSYNNMGADSGSTFSLVERALRETF 301
Oy 293 RYKRSLOLHYRALYTSEQEMHDMIRGELASGRPYUYAGNNOISGHAFVCDYASDGTFFHN 352
Db 302 HYKKSRLYHRSLLPKKEMKDMIRKELAEHNPYUYAGDQSMGHAFVCDGEPDGFHN 361
Oy 353 WGGGYSNGYKTLTLLSPSLGIGEGIGFTTYOEITITGTEPAKTAENACTALPLALK 412
Db 362 WGGGSGNSGFYLNLLNPGSLGAGDGGYSTDOEVVIGIEPASNEA--PGIVDPPTTLY 420
Oy 413 DIFAEKXSESGLVNGYSIYNTGEGSNLDLGYRLNKADGEVIEVKTSSINISWYG--YGEH 471
Db 421 GLO--HMSDAILLSKIKKYSTYAGDVKATLRLTPNGETTNPAVTVPTVEMDIIGES 479
Oy 472 PESFSLAPNQLSGINTTLLRYRTGTQOMEPRYHAGGYVNSIKYNTDPPNNVYTVDN 531
Db 480 TGNITIPCSOFAGKNTISILYRTDGMADKELKILMLGVNKEV--TMPAGDVAVSY-- 536
Oy 532 NECKLSIVPNSFYADLNSYHSHITTYOENSDSPDEIRTPAFALSTGATADVYISLGVW 591
Db 537 ADRIYALKDGLSHNKAAYSCKLSATVYVNPGEERSRPTFALRN--TEGRLYTFGRHL 594
Oy 592 AEV--PGSSSYPYVMSKDVLTLSGDTYTLWY--FSINQKDEMKKIGSVKPTPEYTH 648
Db 595 VELHPDDEDEKSLTITGAKARAGYMLVCTGDMELMADSMIELASIEVAEHSTHS 654
Oy 649 PLFEVGHNGTSTYTLDMANRVLDPFTLKNL--GLPENGELVYVROTQSSGSLMAOET 707
Db 655 SLIVASNPOIDLTLVHRANPELPTFSITNEGATFSKIEIVAKFSET--FFQAKKEH 713
Oy 708 VHKQGETFYKRVVEGP-----IPGSTRALHAFVNGOOLY-----LKKRNTY 754
Db 714 MSLAQGETKVLSEPLANSSLYTNALFPDGIYIVIR-----EOGFMPIDLFGDYR 768
Oy 755 VKIYNGTAAVEIASSE--EIRVPNPARDVEISAPCIQETSIILFDSGLKIVMNSL- 811
Db 769 IRLITLSSDDIAGKDVSTIVLYPNRAHDIYHVAIPTTAGSTLRLFDIOGRMQLSTKIR 828
Oy 812 -----SAGH 815
Db 829 ICRYASRRRTSEGH 843

```

RESULT 2

09AJD2 PRELIMINARY; PRT; 398 AA.

AC 09AJD2; 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)

DE Pyrogenic exotoxin B.

GN SPE B.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1314;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN=N2131;

RA Watanabe Y., Ohkuni H.;

RT Cloning and expression analysis of Streptococcal pyrogenic exotoxin B

```

RT (SPE-B)/Streptococcal cysteine protease (SCP).";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB051298; BAB40954.1; -.
DR InterPro: IPR000200; Peptidase_C10.
DR Pfam: PF01640; Peptidase_C10; 1.
DR PRINTS: PR00797; STREPTOPAIN.
SQ
Sequence 398 AA; 43218 MW; 07D1F534E7867CDD CRC64;

Query Match 11.2%; Score 498; DB 2; Length 398;
Best local similarity 30.6%; Pred. No. 1,9e-21;
Matches 129; Conservative 74; Mismatches 154; Indels 64; Gaps 13;

Oy 6 LIAIVMLFGITAMG-----HSAPVTERKALSLARLALROVSLRMGOTAVSKISIDYV 59
Db 10 LLSLILGFEVLNLPYFADQNFARNKEKADSAITFLOKSAIKAGARSAED--IKLDKYN 68
Oy 60 ROGDAERGITSOEGSPAYFYVANRGNNGYALVADDDHPTPLAVSPIGRPMDSMPN 119
Db 69 LGGEL-----SGSNMTYVNIISG---GFVIVSGDKSPELLGISTGSDVDNG--KEN 116
Oy 120 LRMMLOIY-----DOEIGLILSGKAQLENEELRTEGVAEYHALMDNGHFANDPMNNG 173
Db 117 IASFMSYVEQIKENKKLDITTYAGTAELIQPV-----VKSILD-----SKGIHYNQ 162
Oy 174 GYFNNKKEPILPN-----GNHAYTGCVATAAQMIRYHWPLOGESSFEDY----- 218
Db 163 GNPYNLTPYIEKRVKGEOSFYQHAATGCVATATQIKYHYPKKGKLDYTYLSSNN 222
Oy 219 ----HAGSLVGNMSTGFEYMDWIN--MPGNPLDNLTOSOVAYATLMMDVSAVSMSPY 273
Db 223 PYRNNKRLN---PALSTQIYNNNLTLPYSGRESVQKM--ALSELADVIGISYMDYG 277
Oy 274 ENGSGTYSYVYVGAALNNFRKRSLOLHYRALYTSEQEMHDMIRGELASGRPYUYAGNNS 333
Db 278 PSSGASGRSVQALKENKNGYNSQVHQRNSPDSKQWDSQDKELSQNPYYQGVGVK 337
Oy 334 IGHAFVCDYASDGTFFHNKGGVNGYKTLTLLSPSLGIGEGIGFTTYOEITITGIE 393
Db 338 GGHAFVYDAGDGRNFYHVMWGWGVSDFGRDLALNPALGTGGGAGNGVQSAVVGK 397
Oy 394 P 394
Db 398 P 398

```

RESULT 3

093IQ2 PRELIMINARY; PRT; 344 AA.

AC 093IQ2; 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Cysteine protease SpeB (Fragment).

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1314;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN=SP268;

RA Jadoun J., Eyal O., Sela S.;

RT "Role of cstr, hasa, and speB genes in the internalization of Streptococcus pyogenes by epithelial cells."

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY035886; AAK71464.1; -.
 DR MEROPS: C10.001; -.
 DR InterPro: IPR000200; Peptidase_C10.
 DR Pfam: PF01640; Peptidase_C10; 1.
 DR Protease.
 DR NON_TER 1 1
 FT NON_TER 344 344
 SQ SEQUENCE 344 AA; 37345 MW; FB0A9BD05B541909 CRC64;

	Query Match	4.4%	Score 194	DB 2	Length 424	
	Best Local Similarity	21.6%	Pred. No. 0.0015			
	Matches	96	Conservative	62	Mismatches	161
					Indels	126
					Gaps	15
OY	384	IYEIIITGIEPAKTPAE----	AGTDALPILAKIDEAAYKSSESGLANGYSIYNRGEOS	438		
	:	: : :	:	:	: :	
Dd	82	VLOYIIPLEIFETKNGSRQDQNAOGDESSILEFVDGE-----	LISIG-STG	126		
OY	439	NLDGGRANKADGEYIEVKTSINISWYGCEHPESESFLAPNQLSOGINTTTLTYRRGT	498			
	:	: :	: :			
Dd	127	GIDF-ERINPDIDIEIEVLGA-SSEALYG-----		153		
OY	499	EQWEVVRRAAGGSYSIKVNTTDPNNVVYVDNNEGKLSIYPNSFYADLNSEHSTIYQ	558			
	:	: :	: :			
Dd	154	-----SAIGGVNIITRTAKDPRIYFALNTEGRILYTFGRHLV-----		193		

```

QY      413 DIEAEYKESGSLNV-----GY-----SIYNTGEOSNLDLGRLNADGEVLEK 458
Db      101 DIVYTGKNESDJKINSVELMKNKPGVKSKSVAFSPRANQNMKLYKKEPSSNQDLVDYI 160
QY      459 SSINISWTGYGHPESFSLAPNOLSGCINTITLLYTRGTGEOMEPEVYRHAOGGYVNSIKVN 518
Db      161 AS-----NY-----RSPSGNLNNGGPNFLAFAYTGF-----1866
QY      519 TTDPNNVYTVDNNEGKLSIYNSFADINSYEHSTIYQFNSDSPDEIRPYAFALSTG 578
Db      187 -----NNHKKISFV--DIVYSLNGGCONFKNLLFSODGEKKID--KYDLSTLG 2299
QY      579 ATADDVLSLGVWMAEV-----PGSSSNTPVY-----WS-----K 607
Db      230 STSESMGNHAWPLMGVFEPMKNGQSGKSDIGFLSNFVDNPBEFQMSGPIKYESDSMSFSPK 2899
QY      608 DVLTSBEDYTL-----WYRFST-----NNQKQEMKIGSVSVKPTPEYHPLFEVGHQF 6588
Db      230 IQMLLDEDDNNTINSGESCHNEFMITYSDYDEYSIW-----DIRIYIPKSKSEYKEG 3399
QY      659 STYTLDMAHNRVLPDF--TLKNLIGLPF---NGELVYVERQTOSSGSL--WAAQETVH 7099
Db      340 KTPMTMDLYEAFVLFASYSQETNSGIGYDKNANHYLITVAKKEENGNTLTKYRANANDKIH 3999
QY      710 IKQ--GEFF-----YKPYVEGFIPDGSTRATILAAVNGQOL-----YLGKRN 752

```

Db 400 NKLMSPTFTYTSSANALYTPVD-----INPTKGLVCMSWEYLPCKRI 444
 QY 753 Y--TVKIVNGTAVEAIESS-EIRFPNPARDYVEISAPCIPOETSIILPDLGKIYMN 809
 Db 445 VMSDQWTHANGVEDYIMOGSKMLYENPNPQEVAVISLP--TAANCKRAYVDMQGRVAAE 503
 QY 810 SLASGHRMVSRLPNCAYILKV 832
 Db 504 SFGSNEYRLNVOHLAKGTILKV 526

RESULT 6
 ID 082411 PRELIMINARY; PRT: 3624 AA.
 AC 082411;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 Large repetitive protein.
 STY2875.
 Salmomella typhl.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmomella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Huggill K.L., Bentley S.D., Holden M.T.G., Sebaiha M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltyell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels R.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RA "complete genome sequence of a multiple drug resistant Salmomella
 RT enterica serovar Typhl CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL: AL627276; CAD05867.1;
 DR InterPro: IPR003410; Hyalin.
 KM Complete proteome.
 SO SEQUENCE 3624 AA; 365353 MW; BB52E561D6998F14 CRC64;

Query Match 3.9%; Score 173; DB 16; Length 3624;
 Best Local Similarity 20.3%; Pred. No. 0.64;
 Matches 223; Conservative 102; Mismatches 376; Indels 400; Gaps 54;

14 GIMOGHSAVPTKERALSLARLALROYSLMGOTAVSDKISIDYVROGAERGIR----- 69
 2545 GIA-AGFTAPDTRVPEAPITTVDDVGYITGAINGOVTDNAQPTLNGTAQAGATVSIY 2603

QY 70 -----SOEGSPAYFYVANRGNEGYALVAADRIPTILAYS 106
 Db 2604 NNCALLGTTTANASGNMSPFTPGNLTGESHAFATATNANGTSVSTAATVIYDILAPGT 2663
 QY 107 PIGRFMDSPDLRLMQLQIYDOEIGLISGKAQLEELRT--EGVPAEVHLMONGHF 164
 Db 2664 PSCGTLSDAD-----GGSLSGQAENSTVYVTLGAVTTLTTTA--GSNG-- 2703
 QY 165 ANDPMRNGGYPMNKE-----PLLP---NGNHATGCVARA 198
 Db 2704 -----AMSLTLPTKQIEGOLINTVATDAAGNMSGTIGTAPILPLAARDNITSLDITSA 2758
 QY 199 AAQIMRYHSMPLQEGSGSFYHAGSLVGN-----MSGTFGEM----- 234
 Db 2759 VNSTQVSDYSGLLLVAGL--NVASVLCGNDAQVEFTIAGCGTCDVITIDAATGIVSLIS 2817
 QY 235 -----YD-----WINPQNDLNLTSQSDAYATILMRDVSASVMSFENGSGTYS 281
 Db 2818 TOEIVVQRYDTSIGTY-----TTIVTAVGDFANLLTLTGSGVTILNLGAGGQYR 2868
 QY 282 VVVGALRNFRKRRLQGLHVRALYTS--QEWHDIMRGELASGRPYVYAGN----- 330

Db 2869 V-----LTYNTSLH--ATGSYSLDADVDYHQSAG--IISG--PTISGNVWADDTAPT 2915
 QY 331 -----NOSIGHAFVCDGVAASOSTEPHNMWGVSGEFKLLISTSGIG- 376
 Db 2916 GTTVATATNANGVSTPVGAGV--DILQGTGLHIN-----QGSYTYTLTKFTA--GYGH 2967
 QY 377 GEGIGFTIYQ-----ELTNGIEPAKTPAEA--GTDALPILAKDIEAEY-----KSES 422
 Db 2968 KESFTYITNGVGSSAQLVNLGPAVPVGSVATDNNASLVF--DTHSVYVNGSPISGS 3026
 QY 423 G---LVNGVS-----LYNGEEO-----SNLDL--GYRLN 447
 Db 3027 GTVTLVSLGGLNVLNANLMDMTNPILFNVEEGATRTMTLQGTGVGSVLTSTEDLYYRPN 3086
 QY 448 KA--DDEVIEVKTSSINISWYGEHPESTLAPNOLSGINITTLTTRTGEQPEPR 505
 Db 3087 DAIOYEQFRVQGWINTLLLAGOSPLTTLTGGEYLVNLTASSISVLTGYTLAISOD 3146
 QY 506 HAQGYVNSIKVNT-----TDPNNVVV----- 527
 Db 3147 HTVA--VDSITATNTGNVLNDVAPFDALLTEVNGVAIAATGTEVNGLSLIDARCN 3204
 QY 528 ---TVDNNGKLSI--VPNSFVADLNSYESTITVQFNSDSPDEIRTPVAPALSTGATAD 583
 Db 3205 YTYTLKNGVAGADSIKTPDSFIYTLKAPNGDTDASLINI-----TPARALDAINDVSD 3257
 QY 584 VISL-----GWMAEV-----PGSSNPVYVMSK-----DVLTLSEGD 616
 Db 3258 TISVATLODTAAMLDSSVGSAGKLGKSGSGSTEDVATGYLKASLAFVDSVL----- 3313
 QY 617 YTLWRFSSINNOKEMK-----KIGSVYKTPREYTHPLFEVGHNOTSYTLDMAN 668
 Db 3314 -----ITLGNLMSAIDENGVIINGVTPVANIT-----LG--SATYVNLSS-- 3354
 QY 669 RVLPTDTLNLGLPFL--GELVYVFRQTOSSSGSLMAOE-----YTHIKGEFTVYKPYVE 723
 Db 3355 -----GLELDAGTYTLNFTGTNLTAGATITTPHYIGTVDLDFEFTSGTHTVL- 3402
 QY 724 GPIDGS-----YRATLHAFVN-----GOOQLTKGK 750
 Db 3403 GNIFDSDDAAGAMDQANTVNTRLISIGYNSAATLDAANNTSATIGHGTLQIHLDA 3462
 QY 751 RNTYKIVNGTAVEAIESSSEIRFPNPARDYVEISAPCIPOETSIILPDLGKIYMN 810
 Db 3463 YTYTLN--NGVAMSSITSKE--VFYQLDDKRIGHT-----DSATLTIDMAPQIVSTNQ 3511
 QY 811 ---LSAGHRMVSRLPNCGA 827
 Db 3512 NDVLIGSAYGDITLHYHLNGA 3532

RESULT 7
 ID 093417 PRELIMINARY; PRT: 1330 AA.
 AC 093417;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE MS115, putative beta-agarase.
 OS Micrococcilla sp. PREL.
 OG Plasmid pSD15.
 OC Bacteria; CFB group; Flexibacter group; Micrococcilla.
 OX NCBI_TaxID=155537;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PREL;
 RA Zhong Z., Toudkardian A., Helinski D., Knaut V., Sykes S.,
 RA Wilkinson J.E., O'Byrne C., Shea T., De Loughery C., Caspi R.;
 RT "Sequence Analysis of a 101 kb Plasmid Required for Agar-degradation
 of a Micrococcilla strain";
 RL Submitted (JAN-2001) to the EMBL/Genbank/DDBI databases.
 DR EMBL: AF339846; AAK62837.1; -.

DR InterPro: IPR003343; Big_2.
 DR InterPro: IPR000847; HTH_LysR.
 DR Pfam: PF02368; Big_2; 3
 DR PROSITE: PS00044; HTH_LysR_FAMILY; UNKNOWN_1.
 KW Plasmid.
 SQ SEQUENCE 1330 AA; 143942 MW; 083073D8BFCB28F CRC64;

Query Match 3.7%; Score 166; DB 2; Length 1330;
 Best Local Similarity 20.0%; Pred. No. 0.37;
 Matches 110; Conservative 101; Mismatches 337; Indels 244; Gaps 37;

QY 91 ALVAADDRPTLLAYSPIGRFDMSPDNLRLMQLQYDDEIGLLSGAQLNEILLTEG 150
 DB 614 ALYNDND-----NOVGHLSIDKLTVNCT-----EITPLDMGTNGVRRRYFSTLE 659
 QY 151 VPEVYALNDNGHFADNPRMNOGYPRNNKEPLLPGNNAHYTCGVATAAQIMRYHSWPL 210
 DB 660 IVPVGLLTQNNFTVD-----FHHVEVAVVNLQTEFEFSKVP 658
 QY 211 QGEGSPDYHAGSLVGNMGTGFGAMYDWINMGNPDNLNQSODVADATLMDVSVSM 270
 DB 699 RSTSSDPVAVTGVTS-----PGSPYAQGSTVQMTAHVQPTNATDOSVYW 744
 QY 271 SFYE-----NGSGTYSVYVVG-----ALRNNFRKRSLOLHY-----R 303
 DB 745 SSNNANVAVSNAGCEVTGIAQGTATITATNDGFIASQVWVTTGDVDTGVSVPTSA 804
 QY 304 ALTTSQ--EWMHMRGELASGRPVYAGNNOI-----GHAFCVDCVATSDGTFHNMWCG 357
 DB 805 SILVGSIDLTETVSPFNATDKSVTWTTSAAVTVNGSGLYAKNGSAVTVTTNDG 864
 QY 358 VSNQFKLLTLPSTLIGEGIGFTIYOEITGI-----EPAKTPA-----EAGTDALP 407
 DB 865 -----FTQSAVYTTTGGGSAIYEAESFTSTGTTDDSPYGGGIGVNNAGIN--- 914
 QY 408 ILALKIEAEYKESGLNVG---YSI-YVTGEEQSLDLGRLNKAAGEVIEVTKSSIN 462
 DB 915 ---INVYSEDMAEYGINVSEAGTYQIEYQISTPSNNAQVRFEL---DGNVSTDPNPN 968
 QY 463 ISMYGGEHESFSLAPNLSOGINTTLTYRRGTQEMEPVHAOGGYVNSIKVNTDP 522
 DB 969 GQMDSTKLIASTI--STLSTGSHYVRLVAGANAMQW-----NLKRYLTXT 1015
 QY 523 NNWVTVVNDNEGKLSIYVNSFVADLNSYESTITVQFNSDSPDEIRTPVAFALSTGATD 582
 DB 1016 GS-----STVAVTGVSAVPTNV-----SLISGSTD 1041
 QY 583 DVISLGMVAEVPGGSSNTPVWMSKD-----VLITSEGDYTL 619
 DB 1042 LLETIVN-----PGNATDKSVMSNNTAVATVANGLVAGAGTAVITVTSDDGHTA 1095
 QY 620 WYFESINNOXDEKKIGSVKPTETHTHLEFY-----GHNOTSYTTLDMANRY 670
 DB 1096 TCSVTVYSGNSVELTQADEPAT--TGCTHDGFQVYVNGVTAATNMOTGMA---DYSVT 1151
 QY 671 LP---DETLKN-LGLFENGELVYVFROTOSSSGLMAAOETVHI-KOGEFYVKKPYVEGP 725
 DB 1152 IIPAGGYSTIEFYNGTIVNGAAVTI-----SYDG---AVQRTDAVPNNGNMDFVESLAVG 1203
 QY 726 IPDGSRAFLHAFVNGOQOYLKGRNY-----TVKIYNGTAVEAIESS-----EE 771
 DB 1204 -----RISLNA---GSHTRILGDDGTNGEMEMMDRVLKSGAASSVESSSQIYNQG 1254
 QY 772 IYFPPARADYVETIS--APCIFQETSIILFDLSGKIYKNSLSAGHR--MDVSRPLNGAY 828
 DB 1255 ISITVPPADKRIYVRLAPDLQULT--ISNVSGKIYRKMSVEGPNDYILDVGLKTVGY 1311
 QY 829 ILKVDGYTTKIN 840
 DB 1312 FLHFRGSKTYFN 1323
 RESULT 8

093416 ID 093416 PRELIMINARY; PRT; 614 AA.
 AC 093416
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MS116, putative beta-galactosidase.
 OS Microscilla sp. PREL.
 OC Plasmid PSD15.
 OC Bacteria; CFB group; Flexibacter group; Microscilla.
 NCBI_TaxID=155537;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PREL:
 RA Zhong Z., Toukardarian A., Hellinski D., Knauf V., Sykes S.,
 RA Wilkinson J.E., O'Brien C., Shea T., De Loughery C., Caspi R.,
 RT "Sequence Analysis of a 101 kb Plasmid Required for Agar-degradation
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF339846; AAK62838.1; -
 KW Plasmid.
 SQ SEQUENCE 614 AA; 67878 MW; A9571EC7D32A7793 CRC64;

Query Match 3.7%; Score 164.5; DB 2; Length 614;
 Best Local Similarity 19.4%; Pred. No. 0.14;
 Matches 138; Conservative 82; Mismatches 262; Indels 229; Gaps 32;

QY 222 SLVGNMGTG--EYMDIMNPNPDNLNQSODVADATLMDVSVSMSTYENGSG--T 279
 DB 22 TLITWTAFAFGAARODSGIP-----VPASA-----GCGKT 52
 QY 280 YSYVYVGAALNNFRKRSLOLHYRALTYSQEMHDMIRGELASGRPVYAGNNOISGAFV 339
 DB 53 WQLOASADPFNTFNETSOL---TNFGSKMWN-----FY 85
 QY 340 CDGYASDGTPTFMWGMGVSNGEYKLTLL---SPTSLIGEGIGFTIYOEITGIEPAK 386
 DB 86 HNGMDGEGTYYQYNHVSVSGNLVLRASNPSTTKMGVPGVANG---CITSNNRVK 139
 QY 397 TP--AEAGTALPLIALKDI-----EAETSEGLNVGY-----STY 431
 DB 140 YPVEFVSAVSVAINALASDVWLLSPDDTDEIDIECYGAGSNNAYPEAFIHLSHSFEV 199
 QY 432 NTGEGQSLDLGRLNKA-----DOEVEYK 457
 DB 200 NPFODIOPRLNLSMKSGSVSMGDIYCMNNKRYVAVGNWGPKEFYIIDGELRYVL 259
 QY 458 -----TSSINISWY-----GYGEHESFSLAPNLSOGINTTLTYLRRTG 497
 DB 260 YDKAFATKVGCTWYTYPTMTNGSLDRSGY---QSVQYATGSSYSFSTLQASNTSS 315
 QY 498 TEOMEVPRHAOGGYVSIKVTNTDPNNVYVVDNNEKLSIVNSFVADLNSYESTITV 557
 DB 316 TSVIIDLPRNY--QGGN-----GFTKELDIITINVEQNMHVRAGRTPSADLNDPAKMKV 368
 QY 558 QPNSDSPDEIRT--PVAFALSTGAT-----ADVDISLGMVAEVPGGSSNTPVWMSKV 609
 DB 369 -----DWIRYKPYTTGTGGGGSATYEAESFATGTGYNDASSGGGTYGKAGCTI 421
 QY 610 LITSEGDYTLMYFESINNOXDEKKIGSVKPTETHTHLEFYGHNOTSYTTLDMANR 669
 DB 422 INVYNG--CWMVEYPI-----VPEE--GYVEGYH---YATPISNTS 457
 QY 670 VLDPFTLKNIGLPPNGELVYVFROTOSSSGLMAAOETVHI-KOGEFYVKKPYVEGP 729
 DB 458 V--QFSYS--GTFP-----FTTTLTNTGG--MGNTQTRMASOTANPT-----AG 495
 QY 730 SYRATLHAFVNGOQ-----QYLKGRNRTVKNCTAVEAIESSSEIRVPPARADY 783
 DB 496 NHLIKLTAGSADQMNQKTYTLTRVGNMNSRYLIVPTSDVEGY---DIVFPPPTDDL 551
 QY 784 EISAPCIPOETSIILFDLSGKIYKNSLSAGHRMDVSRPLNGAYILKVG 834

Db 552 RI0G-LDDKMYVYLYDLGNTNYSRQYKGAERLADVSQLSDGILYLLKLEG 601

RESULT 9

08KX8 PRELIMINARY; PRT; 1588 AA.

ID 08KX8

AC 08KX8

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Stalidase-like protein.

GN CpeI254.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Clostridiaceae; Clostridium.

OX NCBI_Taxid=1502;

FN [1]

SEQUENCE FROM N.A.

STRAIN-13 / TYPE A;

PUBMED-11792842;

Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A., Shiba T., Ogasawara N., Hattori M., Kubara S., Hayashi H.: "Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater."; Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

RT EMBL; AP003189; BAB80970.1; -

DR InterPro; IPR000421; FAS8_C.

DR Pfam; PF00041; fn3; 1.

DR SMART; SM00060; FN3; 1.

KW Complete proteome.

SQ SEQUENCE 1588 AA; 178194 MW; D34C534B41E04D49 CRC64;

Query Match 3.7%; Score 164; DB 16; Length 1588;

Best Local Similarity 18.0%; Pred. No. 0.63;

Matches 206; Conservative 164; Mismatches 361; Indels 412; Gaps 58;

QY 2 KKSFLAIYVLFGLAMOGHSAPYTKERALSRLARLRQVSLRNGQAVSDKISI--DYVY 59

DB 8 KKLISLTL-----GLSVALTNTPALSLVLANGAKVEDVYVASKS--QDKITIGNEYIS 56

QY 60 RQGDABERG--ITSQEGSPAYFYVANRGNNEGVAL--VAADDRIPIILLASPIGRDMS 115

DB 57 RESEIENGKVLTEISEIENSRAHTLVLPOLGSEDTITNTIOENSDLPVEANSP----- 108

QY 116 MPDNLRMWLOIYDOETIGLILSGRAQLNEELIRTEGV--PA-EVHALMDNGHFANDPMRN 172

DB 109 -----KEVLD-----RANMNAITLANSSTAYATDEKLLDDGKNTYIDSYNI 151

QY 173 QGYPMN-----NKEPLLPRGNHAYTGCYATTAQAQIMRY-----HWPILQ 211

DB 152 TGIPTSLKIDLGKVTSSFSYQKRPFTDANGKNGTM---GQYKLYVEDGEMWTEA 207

QY 212 GEGSPD-----YHAGSLV--GNNSGFEGEMV----- 235

DB 208 GESEFREDENLHQEENLHNVGVYGNDKYTESKIRIDQLSDALGTOERSGAEIYL 267

QY 236 ---DWIMPGNP-----DLDNLTQSQVADATLMDV---SASVSNFENGSGTYSY 282

DB 268 YSDKYKEPEEPIAPKAIKSSDLT---IDTSFKIEDINGKKLITISYEPYONGEYNI 324

QY 283 YVVGALRNFRYKRSQQLHVR-----ALYISQEHMDMI--- 315

DB 325 DWYTVLENDHWRSP-LEIKTNENNAKIDYIDLHFILEDGIKDTLWHPLEDTYITL 383

QY 316 --RGELASGPYVYAGNNSIGHAF-----VCDG-----YASNG 347

DB 384 IKNELMLGQPIY--ANGMFSESEPAATDYYDDGMQRIYSGSFEELEDNOLTTDG 441

QY 348 TTHFWN-GWGVSNCF-----YKLTLLSPIS-----LGIGGEGIGFT 383

DB 442 KF-VSMQNVVGAAGKGVDTDVQTDFFEYISEIATPTDFRKQYNSWYDNMLITIDSISKS 500

QY 384 IY--QELIT--GLEPAK-----TREAG-----TDALP- 407

DB 501 FYGTGKGLTENGVEPIIDSYVDDGMHNYRDPPEPNPISKEQVGNMRTGFEFNKDFPN 560

QY 408 -IILAKDIEAEYKESGL-----NVGSIY-----NTGEQSNLDLGRNLKNADE 452

DB 561 ELTSTSELNISKPSKSGMLNMGPGGYNFNGGFARILEKMGTYAQTN--NGNVVCGSR 618

QY 453 VIEWKTS-----SINISWYGYGEHPESFSLAP-----NOLSGQINTI--TLYRRT 496

DB 619 YKNTLSLFLDYOKRFDIDYWKL---DGFALPCTSENNHDMHTGCHNNMYTTDLMEK- 673

QY 497 GTEQWEPVRRAQGYVNSIKVNTTDPNNV-----VYTDNNGKLSIYPSNVADL 547

DB 674 WTPAMETMRASRAEKKOLFYNATCYVNSPWLQVNTVWQNSO-----DTGEAGT 726

QY 548 NTEHSTIT---VOFNSDSPDEIRTPV-----AFALSTGATADVISLQWVAEYP 595

DB 727 GSRHQOKITYRDVAVHYDIKSNQIOFPANITYNHEPIYGVSDGFATTEDFDPLFANV 786

QY 596 GGSNNTPVWS-----KDVLLISEGDTYLTMTFSTNNOK-----DEWK-K 634

DB 787 RGTAFELWLYSPSINMDEKMKVNAVLDVENNFPVLEKAKLFGHRAEGYVYGSAMGN 846

QY 635 IGSVSVKPTPEYTHPLFEVGHNOTSYTIDMAH-----NRVLPDETLKNTGL 681

DB 847 EGVTSFRNPT-----GETKEYITDLDIYGVKPSVNLGNQVLP-YKVGDDGS 894

QY 682 PRNGELVYVFPROTOSSSGLMAAQETVHIKQGETFYKPVGPIPDGSYRATLAFVNG 741

DB 895 VSYGDSITVTL-----PYETRILOYGKV-----DNKACEIVSAKVTG 932

QY 742 QOOLYKGRKNTVKTNGTAVAEIESP-----EIVFPNP 778

DB 933 ENETITK---INERENEEYVSGENITETFKLLDRTVYISTSKNLOEAKLNING 988

QY 779 ARDYVEISAPCIPOETSI-ILFDLSGKIYMKNSLASHGRMDVSRPLNGAYILKVDGYT 837

DB 989 EBDALQ-----NPLTSLTIPYNNKGIISVSGELVGENINKKYN-----NSDYFL 1039

QY 838 KIN 840

DB 1040 ELN 1042

RESULT 10

058048 PRELIMINARY; PRT; 1155 AA.

ID 058048

AC 058048

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE 1155AA long hypothetical PYROLYSIN.

GN PH0310.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;

OC Pyrococcus.

OX NCBI_Taxid=53953;

FN [1]

SEQUENCE FROM N.A.

STRAIN-OT3;

MEDLINE-98344137; PubMed-9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Hatakeyama A., Nagai Y., Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamaya M., Ohnuku Y., Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Shizuya H., Nakamura Y., Robb F.T., Horikoshi K., Masuchi Y., Shizuya H., Kikuchi H.;

RT "Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";

RT EMBL; AP000001; BAA29383.1; -

DR InterPro; IPR000668; Peptidase_C1.

Best Local Similarity 17.7%; Pred. No. 0.6;
Matches 146; Conservative 111; Mismatches 242; Indels 327; Gaps 37;

```

OY 221 GSLVGMSTGFGEMMYMINP---GNPDLNLTQSOYDAVATLMRVASVSSEFENG 277
DB 219 GONVGLTNNYPEDY---QPEKGNBEL-QLSGVITVANGTLLAOIK---DVSLLEYD 270
OY 278 GTSYVYVVALRNNFRYKRSLOLH---VRALYTSQEHMDMREGELSGRPYVYAGNQSIC 335
DB 271 STVS-----LRDN-----PLQVEYCVTAAY-----DESTE 295
OY 336 HAFVCD---GASGCTHFMGMGAGVNGEYKLLSPTSIGIGEGIGTITVOEITTCIE 393
DB 296 SSTVCGTLLHAYTDALILENFENCPVNGW-----LVTDADDGFSWGHYL----- 340
OY 394 PAKTPAAGTADLPILALKDIEAEYKESGLANGY-----SIYNTGEOSNLDCGRFLNK 449
DB 341 -----NAYDAFP-----GHNGHCSLSASYIPGIGPTPPDNYLTPKY 378
OY 450 DG-EVIEVKTSSINISWYGEHPESESLAPNOLSGINTITLLYRRTGEQ---W-E 502
DB 379 EGAKRYKVMYSTODAMMAA-----EHYAVMASTTGAVGDEVILFETMTAKPTGAMYER 433
OY 503 PVRHAOG-GYVNSIKVNTD-----PNNVVTYDNNEGKLS-I 538
DB 434 TTNLPBGTYIAMRHYNCTDIYFLKLDITVEGTPASEPEPVDFVSLIENKGRKLN 493
OY 539 VPNSF-----VADLNSYESTITVOFNSDPEIRTPVAFALSTGATADVIS 586
DB 494 YRNGYEPDKTDKDLQLAGYNITANGSLVHIQD-----TVLEYIDETYSRDDOYE 547
OY 587 L-----GMYAEVPGSSNYPV 604
DB 548 VECVATVYNDNIESQVCDKLIYDSQSDIILEGFEAGISPEGMILLIDADGNVNM--- 604
OY 605 WSKDVLTLESGDTLRYRESINNOKEMKIGS-----VSATPEY-THLPFEG- 654
DB 605 -----DYPTPTMG---HDEKCIASPSYLPIMIGVLTPTNYLTPLLEAKIV 649
OY 655 ---HNOTSTYTLDMANRVLDPDTLKNLGLPENGELVVFROTQSSSGLMAAQETVH 710
DB 650 KYWVSQADAVYSME-HYAVMST---GTAVE-DFVLLFEETMAKANGAMYEKTIYL 702
OY 711 KOGERT-----YKRVVEGPIPDSDYRATL----- 735
DB 703 PAKTYIAMRHYDCTDMFELLDDITVYRSTETVPEPVDFVSLIENKGRKLMYPNG 762
DB 736 -----HAFVNGQOOLYLK----- 748
OY 763 YEPDKTDKKPLQLTGYNITANGSLVHIQDPTVLEYIDETYSRDDQVEMEXCVTAAYN 822
OY 749 -----GKRVYTKIYVGTAVEAIESESEIRVFPNPARDVEISAPCIPOETSIILF 799
DB 823 DNIESQVCDKLTNTI-----TSLDNIOQDTSIKITPNPASYVVRLEG-LSRKSSTIELY 876
OY 800 DLSGKIVMKNLSAGHRMDVSRPLPGAVILKYDG---YTKRINT 841
DB 877 NALGICILREETHSEKTEIDVSRINDGVYLIKRVGKNKTKTEKVEI 922

```

RESULT 15

08THC9 PRELIMINARY; PRT: 2951 AA.
AC 08THC9;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Cell surface protein.
GN MA4588.
OS Methanosarcina acetivorans.
OC Archaea: Euryarchaeota: Methanococci: Methanosarcinales;
OC Methanosarcinaceae: Methanosarcina.
OX NCBI_TaxID=2214;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C2A / ATCC 35395 / DSM 2834;
 RX MBLINB-21929760; PubMed-11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Ahnoot D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grhame D.A., Guss A.M.,
 RA Hedderley R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 DR EMBL: AE011180; AAC07927.1; -;
 KW Complete proteome.
 SQ SEQUENCE 2951 AA; 319558 MW; A630165SD222D25 CRC64;

Query Match

3.6%; Score 158.5; DB 17; Length 2951;

Best Local Similarity 18.7%; Pred. No. 3.4; Indels 379; Gaps 50;

Matches 169; Conservative 111; Mismatches 243;

```

OY 41 SLRMGOTAVSDK-----ISIDYVYRQDAERGITSQEGSPAYFYVNRGNEGYALVAD 96
DB 1988 SLPTITGSDPTIKRFSYTDALQGENELGYT-----GPSYTNFNM----- 2029
OY 97 DRIPTILAVSPIGRFMDSNPDLNRMLOIYDQELIGLISGRAQLNEELIRTE-GVPAEV 155
DB 2030 ---AVLEVYQVYASAEANFTAN-----VTSNAPLNVFEFTDSTGPTGS- 2069
OY 156 HALMDGNHRANDMRNOCGYPMNKRPLRPNGHATGCVATAAAGIMYHSPLOGESS 215
DB 2070 -----WTFDFGDKNSTE---ONPTHTYT-AGTYTVLKVYSNSGSDSEK 2112
OY 216 FDY-HAGSLV-----GNMSTGFE----- 233
DB 2113 TGYITAGSVYLAIVANFVSQDTGTAPLSVQFDESTNPTMTWTFEGDKSTEQGNPH 2172
OY 234 MDWT-----NMPGNPDLDNLTOSQVDAVATLMDVSAVSASTYENGSGTYSVY 283
DB 2173 TYETIGTYVKLTAIYVGS---NFT-IKTD-YITVTSVSAVASFEDNSG----- 2221
OY 284 VVGLANRNRFRYKSLQILHVRALTYSOEHMDMIRGELSGRPYVYAGNQSIGHAFVCDGY 343
DB 2222 -----RVPTVO-FTDITGVSYS----- 2239
OY 344 ASDGTFHFMW---GMGVSNN-----GFYKLLSPTSIGIGEGIGTITVOEIT 390
DB 2240 -----WMMDFGDSNBNQNTHTYVTEGVSNTVL---TANGPG-----SNTIT 2281
OY 391 GIEPAPTPADAGTDA---LPIALAKD-----IEAEKSEGLANGYS----- 429
DB 2282 STEPVVVSAPLTSDSNGGIPLTIVONGVSGDLMDYSYAMTSAQKAFITLPSYDIKW 2341
OY 430 -----IYNTGEOSNLDCGRFLN-----KADG-----EYIEVTSINISWYGEHPES 474
DB 2342 ARLYVDVYD-GHEKN---YKGNVEISIDADDSTYELQKNTFTTYSFPEG----- 2391
OY 475 FSLAPNOLSGINTITLLYRRTGEOMEVVRHAOGVNSIKVNTDPNNVVTVVNN-E 533
DB 2392 -GTGPVWLDHLNRYTSQD---QMMYDL-----TGEISQGTNVVQAITSKIDSND 2438
OY 534 GKLSIYVNSFVADLNSYESTITVOFNSDPEIR-----TPVAFALSTGATA-- 581
DB 2439 GRVKAM-----TLVVAYDDSDSEVYVWVNOGHDTVPDLTEY-TGSTSG 2483
OY 582 DDVYSIGWMAEV-----PGSSNMPVYVMSDVLTL-LSGEGY 617
DB 2484 TSTLASGWSANLTAIYLASVDGIYSFOGTTLTSGTPGSGSYGDNTW--DVSSMLTAGET 2541

```

```

OY 618 TLWRFSTJNNOKDEWKKT--GSVSYK-----TPTEYTHBLEFGHNOQTSTYLLDMAHNRVL 671
Db 2542 SI-----FTLNKQDEKTYKKIPLMALMSYKYKAGSPTA-----PLAGFSANVTGEYV- 2586
OY 672 PDLTKLNIGLPENGELVVFPRQTOSSSGSLMAOETVHIKQGEFFVYKPVVGGPIPDGSY 731
Db 2587 -----PLTVLFSDESIQSPITAWMD-----FGDNET-----SSE 2615
OY 732 RATLHAFFNGOQQLXKGRNRYTVKYNGCTAVEALIESESEELRVFENPNPARDYVEISAPCIP 791
Db 2616 QSPHPTY-----SAGNRYTVTL-----TVTNAAGSDSEVKT-----DYIISESSEMP 2657
OY 792 QE 793
Db 2658 EE 2659

```

RESULT 16

08YKJ3 PRELIMINARY; PRJ: 4936 AA.
08YKJ3.
01-MAR-2002 (TREMblrel. 20, Created)
01-MAR-2002 (TREMblrel. 20, Last sequence update)
01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical protein ALR7304.
GN ALR7304.
OS Anabaena sp. (strain PCC 7120).
OC Plasmid PCC7120alpha.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RX Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RX Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RX Kishida Y., Kohra M., Matsumoto M., Matsuno A., Muraiki A.,
RX Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RX Yasuda M., Tabata S.;
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL, AP003600; BAB78386.1; -
DR InterPro: IPR003644; Calx_beta.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001343; Hemlysn_Ca_bind.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00316; Calx_beta; 5.
DR Pfam: PF00353; hemolysinCabin; 25.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PRO0313; CABNDNGRP.
DR PRINTS: PRO0723; SUBTILISIN.
DR SMART: SM00237; Calx_beta; 6.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00330; HEMOLYSIN_CALCIUM; 8.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR Plasmid: Hypothetical protein; Complete proteome.
SQ SEQUENCE 4936 AA, 519407 MW, FBSA8323CB29CG28 CRG64;

Query Match	3.5%	Score 157.5	DB 16	Length 4936
Best Local Similarly	21.6%	Pred. No. 8.2		
Matches 153	Conservative 108	Mismatches 263	Indels 191	Gaps 40

QY 230 TFGSMYMINRGNPDLNLTQSOVDYATILMROVSAVSMSFEENGSGTYSV--YYVGA 287
 Db 58 SFDGVNDVYVNGAKSGLEVSTDTITTEAMINPTGSGSSTIEGCIIVNEEGEEVAFRSDGT 117
 QY 288 LRNNFRYKRSIQLVHRYALYTS--OEW-HDMIRGELASGRPPYYAGNNOISGHAAVCDSYA 344
 Db 118 IRMAFANNNPWTWIMINTSYAFPLQWQHIAITYELG----YIKRYSNVLVHLYT--NMSG 171
 QY 345 SDGFFHFRHWGCVSNGEYKLTLLSPSSLGICGEGCTTYIOEIIYGT---EPAKTPAEA 401

```

Db 172 NIDGFPHAN-----EDDFRIGRIGNOLFGOSIDVIRIWNKARTQAEI 214
Qy 402 GTALPLIALKIOLEAEYKESGLNNGYSIYNTGEOQSLUDGTYRLKADGEYIE-----455
Db 215 QABLIRELTGK-----ETGL-IGYWFNFSINGTQVODL--TGONONGNGVLEAQNIG 263
Qy 456 -YKTSI-NISWGYGEHPESFLAPNQLSOSIN-----TTLTY-----RRTGEOMEPY 504
Db 264 IYVTSLSITDYSIY--EPTETINLLTNPFTNANGLCTOKTATALINVDNDAVAAGTFOBNV 320
Qy 505 RHA-----OGGVNSIKVNTDPPNN--VVYTVDNNEKLSIVNSFVADLNSYESTITVOF 559
Db 321 SYALINEGTLVATVLTNRTGESDGAVSATVNLNSG-----SALAASVD-YDNPRPIVNF 372
Qy 560 NSOSPDEIRTPAVALSTATADVDYISLGMVAEYPGSSNTPYVWSKD--VLTLSGDY 617
Db 373 ANGETSKIYV-IPYVNDNPEFNETINLS--LSNPTGGAT--VGTQNTALITIVDDL 425
Qy 618 -----TLWYRSINN--OKDEMKKI-----GSYSVK--TPREYTHPLFEVGHONTST 666
Db 426 POGGTI--NENINNTVNVNGTASINLVRTGSGDSEVSVTLLPSGRTATAGSDYNLRPIT 483
Qy 661 YTLIDMAHNRLLPDLTKNLGLPFNGELVY---VYRQROSS-SGSLMA-----AOET 707
Db 484 VTFANGETSKTINLISQNGGLFFDGDNDVDNPNANSEIKDFTIELMANPATRAASHPET 543
Qy 708 -----VHIKOG-----ETFFYKPVV 722
Db 544 SSGVNAFFNOKVYAIFFPKOGLTGLGTSNDVYAGISIGTNGVTISEHTLTYMPSVLVYNTAL 603
Qy 723 EGRIPGOSTRATILAHVNVNGOOLYLKGR-----RNTVY---KIYNGTAVEALE-----767
Db 604 SG-----NMLHALVYENKTPKLYINGOFIKAGLTSOYIVPSSLSFGTSTIRODMSFKG 657
Qy 768 SSEIRVFPNPARDYVEISA-----PCIPQETSIILF-----BLSGYVAKNSLSAGH 816
Db 658 SIDDVYKIV-HKARTEEIKAGLNRELITGNESSLICIYWFNFSINGNITVODLSLNKNNNG 713

```

RESULT 17

ID	Q8R8S9	PRELIMINARY;	PRT;	948 AA.
AC	Q8R8S9;			
DT	01-JUN-2002 (TrEMBLrel. 21, Created)			
DT	01-JUN-2002 (TrEMBLrel. 21, last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, last annotation update)			
DE	Intimin.			
GN	EAE-THETA.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia			
OX	NCBI_TaxID=562;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-11201;			
RC				
RA	Schmidt H., Oswald E., Koehler B.;			
RL	"Phylogenetic analysis of the eae gene family ";			
RL	Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AJ308550; CAC81930.1; ..			
DR	SEQUENCE 948 AA; 103292 MW; 895F87D118D6C06B CRC64;			

Query Match	3.58;	Score 156;	DB 2;	Length 948;
Best Local Similarity	19.18;	Pred. No. 0.87;		
Matches 185;	Conservative 141;	Mismatches 330;	Indels 312;	Gaps 50

QY 9 IVALFCIAOGSHAPPTK-----BEALSLALALROYLSLEMGGTAVSDKSID 56
 Db 132 LVAAGGVA--GHTNKTATKSPVAAKSNMTDDKALTYA---AQAASIGSOLSRSLNGD 185
 QY 57 YTRIGDAERGITSOEKSPAYFYVANN-----NNEGATLYAADRIP-----T 101
 Db 186 YA---DRTALGIAINGNDAASQLQAWLQHYHTGAELVNLQSGNNNDGSSL---FELLFPYDSEK 239

```

OY 102 ILATSPICRFDMS-----MDNLRLMLQIYDQETG-----LILSG----- 137
DB 240 MLAFGVGARYIDREFRANIGAGORFELDENMIGYVNFIDQFSGDNTIGIGEGYRDY 299
OY 138 -KAOLNEELIRTEG-----VPAEVALMDNGHPANDP-----MRNNOGIPWN- 178
DB 300 FSSSVN-GFPRNSGHESTYKADYDERPANGFDIRFNGILPSPALGARLMEQOYGDVY 358
OY 179 ---NKEPLLPNGNHAYTGCVAATAAAQIMRYSHPNLOGESGFYHAGSLVGN-----WSGT 230
DB 359 ALFNMDKLQSNPGAATVG-----VATPVPILVTMG-VYRHRCT-GNEDLLYSMQ 406
OY 231 FEEDMD--WIMNPGNDLNLQSOVDAYATLMRDVSAS-----V 268
DB 407 FRYOFDKPW-----SQOLEPQYVNELRFLTSGSRYLQVORNNNILEYKKODIL 454
OY 269 SMSFENGSGTYSVYVYVGLRNNFRKRSLOLHVRLYTSQE--WMD-MIRGELASGRV 325
DB 455 SLISPHIDNGTE-----HSTOKIOLNVNSKYGLDRIWDDSLARSO--GGOI 499
OY 326 YVAGNNOISIGHAFVCDGYASDGTFFHNMGMGVSNCFYKLTLLSPSLIGEGIGFTY 385
DB 500 QHSGSGSVQDYAILPAYVGG-----SMYKVTARAYRNRGNSSNNVQLTIT 547
OY 386 -----OEITTEPAKTPAEA-GTDALPILALKDIEAEYKSESGLVGYSIYNTGE 435
DB 548 VLSNGQVQVGVDTFTADTKTSAKADGTETAITYTA-----TYKKNQV 589
OY 436 EOSNDLGLYRLNKADGEVIEVKTSSINISWYGEHPESFSLAPNOLSGINTITLLYR 495
DB 590 AQANPVSFDIYSGDA-TLSARSATIN-----SSGKATVYLLKSK 628
OY 496 TGTGEWEPYRHAOGIVNSIKYNTDPPNNVYTVDNNEGKLS-IVPNSFVADLNSYENST 554
DB 629 PG-----QVYVSAKTAEWTS-ALNA-----NAVIFVQDTRAKSIREKADKTAKADSDAIT 679
OY 555 IYVOFNSOSPDEIRFPVAFALSTGATADVISLGMVMAEVPGGSSNYFPVYKSDVLTLE 614
DB 680 YTVKWKNNKQNPANNSVYFSTNFG-----DL-----GNSNTQI-----VKTQD 719
OY 615 GDYTLMYRFSINNOXDEMKKIGSVS--VKTPEYTHPLPEVGHNOTSTYTLDMANRVL 672
DB 720 GATYKLTGSGVAGNMAIVSAKSEVTEVAPAKFPFVSALIDNN----- 763
OY 673 DDTLKNLGLPFNGELVYVFRQTOSSGSLMAAOETVHIKOGETFYKPVVEGPIDPGSYR 732
DB 764 ---VNIIGTSANGAL-----PNIM-----LRYGO-----FKLAKG--GDGKXQ 797
OY 733 -----ATLHAFVNGOOLYLGKRNVTYKLVNG--TAVEATLESSFEI-----RV 774
DB 798 WRSODPSVASVDA-LGRVTLTKKGTTK--IEVSGDKQTATVYTTINTPKIITVETQNKV 854
OY 775 FPNPARDVVEISAPCIPOETSIILPDLGKIVAKNSLSAGHRMDVSRPLNGAYILKYDG 834
DB 855 IYSVAEATCSTNSGLPSTST-ELKDYVNNKMGAAANSIEYKKNITTAQTOTADKQSG 913
OY 835 YTTKINIV 842
DB 914 WTSTFDIV 921

```

RESULT 18

O8TP22 PRELIMINARY: PRT: 1075 AA.

AC O8TP22;

DT 01-JUN-2002 (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Surface antigen gene.

GN MA1761.

OS Methanosarcina acetivorans.

OC Archaea: Euryarchaeota: Methanococci: Methanosarcinales; Methanosarcinaceae; Methanosarcina.

```

OX NCBI_TaxID-2214;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-C2A / ATCC 35395 / DSM 2834;
RX MEDLINE-21929760; PubMed-11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endlitz M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smitov S., Anoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahme D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umeyam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010852; AAM05167.1; -.
KW complete proteome.
SQ
SEQUENCE 1075 AA; 114778 MW; BBE122C2BA209CAB CRC64;

```

Query Match 3.5%; Score 155; DB 17; Length 1075;
 Best Local Similarity 18.9%; Pred. No. 1.2;
 Matches 202; Conservative 129; Mismatches 349; Indels 390; Gaps 57;

```

OY 1 MKKSFL-----AIVMLGIAMGSHSAPYKKEKALSLARLALQVSLRMQGTAVSDKISID 56
DB 1 MKRIFLICIALDMCKTGDA-----SAVV--ETVEPANTANTTSGPADLVQFTDISTN 54
OY 57 YVYRQDAERGITTQSEEGSPAYFY-----VANRGNNEGYA--LVAADDRIPPI 102
DB 55 ATGWAMDFDGNNSPEQ-NPVHTYSEGTYSLVFSAKNGSDSEKTEYISVLEP- 111
OY 103 IATSPICRFDMS-----MDNLRLMLQIYD-----OEGLIUS 136
DB 112 --LAPANFSADITSGVGPLYVHFTDSSGSDPT--WEMDNDGDVDSTBQNPOTIT 166
OY 137 GRAOLNEELIRTEGVPAEVALMDNGHFANDPNNNGYR-----WNNKEPLLPNGNHAY 191
DB 167 GEGTYSVKLTLYSSAGSDELTQDHTYV-----YPSDFTANTTEGTVPL-TYEF 216
OY 192 TGCVAATAAQIMRYSHPNLOGESGFYHAGSLVGNMSTGFEEMYMIMNPGNDL---D 247
DB 217 TGNVSGLAPPEYNNMAMVDVGDGYDSPNF--NITHYTB-----PGYDVIVAYD 266
OY 248 NITQSOVDAYATLMRDV--SASY-----SMSFENGSGTYSVYVGLRNNFRK 295
DB 267 GMTQVKKAGYITVLEIPVANSADYVTSQMEPLTVNFQDQSTGVSSY----- 314
OY 296 RSLQLHVRALYTSQEMHDIRGELASGRVYVYVAGNNOISIGHAFVCDGYASDGTFFHNMGM 355
DB 315 -----SMDDNDNCTVDS-----TEQNPNSH-----YTSAGNYVNL- 345
OY 336 GGVSNGEFYKLTLLSPSLIGEGIGFTYOEITTEPAKTPAEAAGTADALPILAKD- 413
DB 346 -TVSNA-----GGSD-----SELKTDYITVQAGVAVATNDLISGLVTV 384
OY 414 -IEAEYKSESGLVGYSIYNTGEQ-SNLDLGRLNKADGEVIEYKTSINSISWYGEH 471
DB 385 PASAFAHEETNPVKLVANQNTGATLSNLSIAVYVADVSSGTVPDYTTI-ASLEGDAK- 442
OY 472 PESFSLAPNOLSGINTITLLYRRTGTQWEP-VRAAGGVYNSIKVNTDPPNNVYVVD 530
DB 443 -----TYTTLI-----DPTIYDLEGGYTYTAAV--VDPDNLIAETD 476
OY 531 -NNEGKLSIVPNSFYADLNSYE-----HSTIVQFNSD-----SPDEIRTPVA 572
DB 477 ETNNKSSAAKN--VRYNGYKGIYWEGGSNITTKHTFDLQNLVLTSTPDSAYQPVG 533
OY 573 FALSTGA-TADV-----ISLGWMAEYVGGSSNRYVW-----SKDV 609

```

Db 534 WESRTEWTASDLPEPDGSTEIEKAFLYVAVNM--DQTPGC---YP--WLINENGNTLDN 586
 QY 610 LITSEGGYTLM-----YRSINNQKDEMKKIGSVSKTPEYTHLPLEVEGHN 656
 Db 587 GNISTGCTTYIRDMNSNGAVANTREYGLCYVDYDVKFSAGNSLWTP-----VGEN 637
 QY 657 QTSYV--TLDMAH-----NRVLPDFTLKNLGLP----- 682
 Db 638 KMLPSTLVVYVGNDETERRKQIFINECEDELGSASGYCTPEATATAPFTGMSIDYE 697
 QY 683 -----ENGELY-----VVFROT----- 694
 Db 698 KVTNAMLVSFAGSAPDEGNLLNGNIVATNAMOGSSNGSGPLFEDATINVTGNBAGI 757
 QY 695 -QSSSGSLMAAEVHVKOGETFYKPYVE-----GPVP-----DGSYRATLHA 737
 Db 758 QSTISGADLQQLVVEYESAPADPADTATPTGADPLAVNFTDSTGSPTSMSWD 817
 QY 738 FVNG-----QOOLY-LGKRNVTYKIVNGTAVEAIESESSEIRFEPNPARDYVEISAPCIP 791
 Db 818 FCGDANSTAOPLHTYSAGEYNT---VNLTVENDAGDOFELKT-----DYIAVSE--AS 866
 QY 792 QETSIIIFDLGSKIVMKNSLSAGHRDVSRLPENGAYILKYDGYTTTINI 841
 Db 867 GSTVTLVTFDESSSVSEN--ESTEISYVANSNFPSC-----LSGYNLVTLL 909
 RESULT 19
 ID 0979C6 PRELIMINARY; PRT; 2076 AA.
 AC 0979C6;
 DT 01-OCT-2001 (TREMblrel. 18, Created)
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Hypothetical protein TV1235.
 GN TV1235 OR TVG1271204.
 OS Thermoplasma volcanium.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 CC Thermoplasmataceae; Thermoplasmata.
 CX NCBI_TaxID=50339;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GS1 / DSM 4299 / JCM 9571;
 RX MEDLINE=20570466; PubMed=11121031;
 RA Kawashima T., Amano N., Koike H., Makino S.,
 RA Kawashima T., Yamamoto Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
 RA Nunoshita T., Yamamoto Y., Atamaki H., Makino K., Suzuki M.,
 RT "Archaeal adaptation to higher temperatures revealed by genomic
 sequence of Thermoplasma volcanium."
 Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
 EMBL: AP000995; BAB60377.1;
 InterPro: IPR003674; Cofase-STT3.
 DR Pfam: PF02516; STT3; 1.
 KW Hypothetical protein: Complete proteome.
 SQ SEQUENCE 2076 AA; 228297 MW; AF26CDAFDBE638A5 CRC64;
 Query Match 3.5%; Score 154; DB 17; Length 2076;
 Best Local Similarity 20.8%; Pred. NO.3.7;
 Matches 170; Conservative 116; Mismatches 308; Indels 222; Gaps 46;
 QY 172 NQGYPMNNKEPLPLNGNHAYTCVATAAQIMR--YHSMPLQEGSEFDYHAGSLVGNMSG 229
 Db 1032 SKSPYINNGTILILNSTYNT-----TAKFPLINGISGPIRP---PYVYTA-SYLAN--- 1079
 QY 230 TFEKMTIMNPGNPDLDNLQSOVDAYATLMRDVSAS---VSKSYENG----- 276
 Db 1080 --GELYKVOYLNTTVGSSVYVDLSVKYNOIRADVTYVGKPKVMYVYAEAGASGYSYTNF 1137
 QY 277 -SGYSVYVYVGAIRNFRKRSLOLHVRLALTS-QEMHDMIRGELASRPYVYVGNNSI 334
 Db 1138 VNGYIYVPEGC--NYTVYARSYNTATNTQYVYVNTGMSVYNTITMAAPAVMYVGSASNI 1195
 QY 335 GH-AFVCDGVASDGTFFHFMGMGVSNFYLKLLSPSL-GIGEGIGFTIYQRI--- 389

Db 1196 SRIFVMDSOVNNATVYN-----VSNG--KFSVIVPGVYTIYSGKG-TAYMRTIEIT 1247
 QY 390 --TGIEFAKTPA-EAGDAPILALKDIEAEYKESEGLN-----VG- 427
 Db 1248 ENTOLVNSGSPAVSVLSDS--YVEKAVSSGQYIYVQNGLYGYPYSSNSTHERLPVGI 1305
 QY 428 YSIYNT-----GEOSNLDGYRLNKADGVEIEKTSINISWYGEHPESFLA 478
 Db 1306 YRISTATFIAGLSTAGLVLNLKSNY-----SNTVELSNASIS-----ITALY 1348
 QY 479 PNLOSQINT-----ITLLYRTGTQEWEPVHAAGGYVNSI-KVNTDPNNVYTV 530
 Db 1349 NKRISNENFTVDAVSGVILLXSGNVILLEVPYEVSS--GYENVYPLSTYSPVATVISP 1406
 QY 531 NNEGKLSI-----VPNSFVADLSYEHSTITVOFN-----SSPD 565
 Db 1407 YLEENISLMAKPITLGMKPVYIEASFELYNKS-LKPEFNGYIDLYGLYNTLVNVSQRAY 1465
 QY 566 EIRTPVAFALS-----TGAT--ADDVLSLGVMAVEPGSSNFPVMSKDLT 611
 Db 1466 GYMPGTYIYLSNMGNTNLIYNDAITPSSSNLIDVG-IVANFTSDISGVRLT-TKRGSO 1523
 QY 612 LS-----EGDTLMYRESINNQKDEMKKIGSVKPTETHTPLPEVGHNOTSTYTLDMA 666
 Db 1524 LSPGPVPLIGNYIYV-----AKGSKVSVNEIATNTNIVYNSISLANGYLVNTNSLNV 1577
 QY 667 HNRVLPDFTLKNLPLPENGELY-----VVFROTQSS-----GSLMAOER--- 707
 Db 1578 GNYLVKG--SYFVLPNGSFVLPAGNYYISFNSRFSNYSFYVGNASLYLNRNLTLS 1634
 QY 708 VHIQGETFYKPVVEGPIPDGYSYRATLHAFAVNGOOLYTKGRNYT-----VKIVNGTA 762
 Db 1635 VVMSKELFANTTI-----LNTSPYS---YVVLNSGYSYAGRANSTCVASFMLPTGFT 1667
 QY 763 VEALIESSEIRFEPNPARDYVEISAPCIPQETIILD-----LSGKIYVKN 809
 Db 1688 VYSISTNKKMAF-----NYLIN----PEFTSVFNSLSNSPYVYLTALNCEPYLN 1738
 QY 810 -SLAGHRMDVSR-----LPNGAYILKYDGYTTK 838
 Db 1739 VVISLGSSELLVNSSVNTIYLPNTGYTFGSSISTTR 1774
 RESULT 20
 ID 08TUS8 PRELIMINARY; PRT; 2523 AA.
 AC 08TUS8;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Cell surface protein.
 GN MA3699.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 CC Methanosarcinaceae; Methanosarcina.
 CX NCBI_TaxID=2214;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 RA Allen N., Neylor J., Strange-Thomann N., Deatellano K., Johnson R.,
 RA Linmon L., McEwan P., McKernan K., Jalamas J., Tirrell A., Ye W.,
 RA Zinner R.D., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetter H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayah L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic

RT and physiological diversity.
RL Genome Res. 12:532-542(2002).
DR EMBL/ AE011080; AA07054.1; .
KM Complete proteome.
SQ SEQUENCE 2523 AA: 264288 MW: B3F4DCD961126C5A CRC64;

Query Match 3.48; Score 151.5; DB 17; Length 2523;
Best Local Similarity 19.58; Pred. No. 6.9;
Matches 172; Conservative 84; Mismatches 285; Indels 339; Gaps 43;

```
QY 89 GVALVAADDRIFTLASPIGRFMDSPDLKMLQIYDQIEGLISGKAQLEETLRT 148
DB 1025 GYINLVNDSSGVY---APISEFTTDASSGRVPTVQFTDTTGNVSTMSMDFGDGETSD 1080
QY 149 EGVAPEYHALMDNGHFANDPMRNNOGYPMNKEPL---PGNHAYGCVATAAQAQIMRY 205
DB 1081 EQNP--VHTYVTEG-----THVTLTATGPGGTDATSTMEVAA----- 1117
QY 206 HSWPLOGEGSF-----YHAGSLVGNMSGTFGEYDMINMPGNPDLDNLTSQ 253
DB 1118 ---PL-NTGSYDGGSPLLTVQEGTVSGGLMHGSMWPGSGYGEQTFLLPPTHEI-KMARLY 1172
QY 254 VDAVATLMRDV-SASVSMSEYENGSGTYSTYVVGALBNFRYKSLQHLVRL--YTSOE 310
DB 1173 ADVYCGLMTDIKSFSMTIDIDGGDGly-----ELQKHDAASIDYNTTP 1216
QY 311 -----WHDN-----IRGELASGR-----PYVAGNNOISG 335
DB 1217 AWYNDHVRYTSDYFMWYDLTDNITGNEVNIKATGGSDGRIRKHTLVAYADGDDDET- 1275
QY 336 HAFVCGYASDGTFFHN---KMGVGS-----NGFYKLTLLSPYSLGIG 376
DB 1276 HYVWNOGHDTANTGSGYSTGTSALSDPAVNLSTYCAVNGAYTYNGESLSGGPQ 1335
QY 377 GEGIGF---TIYOEITIGIEPAKTPAEAGTDALPILAKDIEAEKSESG----- 423
DB 1336 GEYFGYNDMNVTEYFNSGEDSTYVYAPTSYKIQIAL--LTASAGSSGGDGPVADFTA 1393
QY 424 -----LNVGYSIYNTGEEQSNL-DLG-----YRLN---KADG 451
DB 1394 NVTSGGIPLVSTFSTDESTGSPASMLWDFDGNSTEQNPVHTYSTEGTYSVNLTVINEDG 1453
QY 452 EVIEVKTSSINISWYGEHPEFSFLAPNOLSG--INTI-----TLIYRR 495
DB 1454 SDELKTDTYTVQAG-----QVATNDLSISGLVNTVPASAVFARETNPVKVLNVQN 1505
QY 496 TGTQOME-----PVRHAGGYVNSIK 516
DB 1506 TGTATLSNISIAVYASDVSGIVPVDTTIIASLAGDAKTTVLLIDPTIRDEGGTVYTA 1565
QY 517 VNTTDBNNVVYVD--NNEGKLSIVNSFYADLNSY-----HSTITVQFN 560
DB 1566 V--VDPDNLIAETDETENNKKSSAKPLRY---NGYKKGIIWEGGSNITRHTFDLQGN 1619
QY 561 ---SDSPDEIRTPYAFALSTGA-TADDY-----ISLGMMAEVPGGSSNP 602
DB 1620 LLYSTQPDAYOPVGMESRTETWTASDLPVPGSTIEKAFLYAYVNW--DOTPG--YF 1674
QY 603 VVW-----SKDVLTSRGDYTLW-----YRSINNOKDEMKKIGSYSVK 641
DB 1675 --WLNINPFGNALDNGNISTGNGTLYRDMNSNGCAVADYEGLCVYVWTDKFSAGNSLVA 1732
QY 642 TPTETVHPLEFVGHNQTSYTLDMAHNRVLPFTLKNLGLPENGELVVVFRCOSSGSL 701
DB 1733 TP-----VGENKNALYP-----STLVVYVGNNDNETRKQI 1761
QY 702 MAAQETVHIKQGETFYKPEVBEPIPDGSTRATLHAFVNG 741
DB 1762 FINEEC-----DELAYSLTGYGTPE---EATAYAPFTG 1792
```

Search completed: May 29, 2003, 13:26:53
Job time : 52 secs